

Modeling the pairwise key distribution scheme in the presence of unreliable links

Osman Yağan and Armand M. Makowski
 Department of Electrical and Computer Engineering
 and the Institute for Systems Research
 University of Maryland, College Park
 College Park, Maryland 20742
 oyagan@umd.edu, armand@isr.umd.edu

Abstract—We investigate the secure connectivity of wireless sensor networks under the pairwise key distribution scheme of Chan et al.. Unlike recent work which was carried out under the assumption of *full visibility*, here we assume a (simplified) communication model where unreliable wireless links are represented as on/off channels. We present conditions on how to scale the model parameters so that the network i) has no secure node which is isolated and ii) is securely connected, both with high probability when the number of sensor nodes becomes large. The results are given in the form of *zero-one laws*, and exhibit significant differences with corresponding results in the full visibility case. Through simulations these zero-one laws are shown to be valid also under a more realistic communication model, i.e., the disk model.

Keywords: Wireless sensor networks, Security, Key predistribution, Random graphs, Connectivity.

I. INTRODUCTION

Wireless sensor networks (WSNs) are distributed collections of sensors with limited capabilities for computations and wireless communications. It is envisioned [1] that WSNs will be used in a wide range of applications areas such as healthcare (e.g. patient monitoring), military operations (e.g., battlefield surveillance) and homes (e.g., home automation and monitoring). These WSNs will often be deployed in hostile environments where communications can be monitored, and nodes are subject to capture and surreptitious use by an adversary. Under such circumstances, cryptographic protection will be needed to ensure secure communications, and to support functions such as sensor-capture detection, key revocation and sensor disabling.

Unfortunately, many security schemes developed for general network environments do not take into account the unique features of WSNs: Public key cryptography is not feasible computationally because of the severe limitations imposed on the physical memory and power consumption of the individual sensors. Traditional key exchange and distribution protocols are based on trusting third parties, and this makes them inadequate for large-scale WSNs whose topologies are unknown prior to deployment. We refer the reader to the papers [6],

[11], [20] for discussions of the security challenges in WSN settings.

Random key predistribution schemes were introduced to address some of these difficulties. The idea of randomly assigning secure keys to sensor nodes prior to network deployment was first introduced by Eschenauer and Gligor [11]. Since then, many competing alternatives to the Eschenauer and Gligor (EG) scheme have been proposed; see [6] for a detailed survey of various key distribution schemes for WSNs. With so many schemes available, a basic question arises as to how they compare with each other. Answering this question passes through a good understanding of the properties and performance of the schemes under consideration, and this can be achieved in a number of ways. The approach we use here considers random graph models naturally induced by a given scheme, and then develops the scaling laws corresponding to desirable network properties, e.g., absence of secure nodes which are isolated, secure connectivity, etc. This is done with the aim of deriving guidelines to *dimension* the scheme, namely adjust its parameters so that these properties occur with high probability as the number of nodes becomes large.

To date, most of the efforts along these lines have been carried out under the assumption of *full visibility* according to which sensor nodes are all within communication range of each other; more on this later: Under this assumption, the EG scheme gives rise to a class of random graphs known as random key graphs; relevant results are available in the references [3], [8], [11], [18], [24]. The q-composite scheme [7], a simple variation of the EG scheme, was investigated by Bloznelis et al. [4] through an appropriate extension of the random key graph model. Recently, Yağan and Makowski have analyzed various random graphs induced by the random pairwise key predistribution scheme of Chan et al. [7]; see the conference papers [25], [26].

To be sure, the full visibility assumption does away with the wireless nature of the communication medium supporting WSNs. In return, this simplification makes it possible to focus on how randomization of the key distribution mechanism alone affects the establishment of a secure network in the best of circumstances, i.e., when there are no link failures. A common criticism of this line of work is that by disregarding the unreliability of the wireless links, the resulting dimensioning

guidelines are likely to be too *optimistic*: In practice nodes will have fewer neighbors since some of the communication links may be impaired. As a result, the desired connectivity properties may not be achieved if dimensioning is done according to results derived under full visibility.

In this paper, in an attempt to go beyond full visibility, we revisit the pairwise key predistribution scheme of Chan et al. [7] under more realistic assumptions that account for the possibility that communication links between nodes may not be available – This could occur due to the presence of physical barriers between nodes or because of harsh environmental conditions severely impairing transmission. To study such situations, we introduce a simple communication model where channels are mutually independent, and are either on or off. An overall system model is then constructed by *intersecting* the random graph model of the pairwise key distribution scheme (under full visibility), with an Erdős-Rényi (ER) graph model [5]. For this new random graph structure, we establish zero-one laws for two basic (and related) graph properties, namely graph connectivity and the absence of isolated nodes, as the model parameters are scaled with the number of users – We identify the critical thresholds and show that they coincide. To the best of our knowledge, these full zero-one laws constitute the first *complete* analysis of a key distribution scheme under *non-full* visibility – Contrast this with the partial results by Yi et al. [28] for the absence of isolated nodes (under additional conditions) when the communication model is the disk model.

Although the communication model considered here may be deemed simplistic, it does permit a complete analysis of the issues of interest, with the results already yielding a number of interesting observations: The obtained zero-one laws differ significantly from the corresponding results in the full visibility case [25]. Thus, the communication model may have a significant impact on the dimensioning of the pairwise distribution algorithm, and this points to the need of possibly reevaluating guidelines developed under the full visibility assumption. Furthermore, simulations suggest that the zero-one laws obtained here for the on/off channel model may still be useful in dimensioning the pairwise scheme under the popular, and more realistic, disk model [12].

We also compare the results established here with well-known zero-one laws for ER graphs [5]. In particular, we show that the connectivity behavior of the model studied here does not in general resemble that of the ER graphs. The picture is somewhat more subtle for the results also imply that if the channel is very poor, the model studied here indeed behaves like an ER graph as far as connectivity is concerned. The comparison with ER graphs is particularly relevant to the analysis of key distribution schemes for WSNs: Indeed, connectivity results for ER graphs have often been used in the dimensioning and evaluation of key distribution schemes, e.g., see the papers by Eschenauer and Gligor [11], Chan et al. [7] and Hwang and Kim [13]. There it is a common practice to assume that the random graph induced by the particular key distribution scheme behaves *like* an ER graph (although it is not strictly speaking an ER graph). As pointed

out by Di Pietro et al. [8] such an assumption is made without any formal justification, and subsequent efforts to confirm its validity have remained limited to this date: The EG scheme has been analyzed by a number of authors [3], [8], [18], [24], and as a result of these efforts it is now known that the ER *assumption* does yield the correct results for both the absence of isolated nodes and connectivity under the assumption of full visibility. On the other hand the recent paper [25] shows that the ER assumption is not valid for the pairwise key distribution of Chan et al. [7]; see Section V-A for details.

The rest of the paper is organized as follows: In Section II, we give precise definitions and implementation details of the pairwise scheme of Chan et al. while Section III is devoted to describing the model of interest. The main results of the paper, namely Theorem 4.1 and Theorem 4.2, are presented in Section IV with an extensive discussion given in Section V. The remaining sections, namely Sections VI through XIII, are devoted to establishing the main results of the paper.

A word on notation and conventions in use: All limiting statements, including asymptotic equivalences, are understood with n going to infinity. The random variables (rvs) under consideration are all defined on the same probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. Probabilistic statements are made with respect to this probability measure \mathbb{P} , and we denote the corresponding expectation operator by \mathbb{E} . Also, we use the notation $=_{st}$ to indicate distributional equality. The indicator function of an event E is denoted by $\mathbf{1}[E]$. For any discrete set S we write $|S|$ for its cardinality. Also, for any pair of events E and F we have

$$\mathbf{1}[E \cup F] = \mathbf{1}[E] + \mathbf{1}[F] - \mathbf{1}[E \cap F]. \quad (1)$$

II. IMPLEMENTING PAIRWISE KEY DISTRIBUTION SCHEMES

Interest in the random pairwise key predistribution scheme of Chan et al. [7] stems from the following advantages over the EG scheme: (i) Even if some nodes are captured, the secrecy of the remaining nodes is *perfectly* preserved; (ii) Unlike earlier schemes, this pairwise scheme enables both node-to-node authentication and quorum-based revocation.

As in the conference papers [25], [26], we parametrize the pairwise key distribution scheme by two positive integers n and K such that $K < n$. There are n nodes, labelled $i = 1, \dots, n$, with unique ids $\text{Id}_1, \dots, \text{Id}_n$. Write $\mathcal{N} := \{1, \dots, n\}$ and set $\mathcal{N}_{-i} := \mathcal{N} - \{i\}$ for each $i = 1, \dots, n$. With node i we associate a subset $\Gamma_{n,i}$ of nodes selected at *random* from \mathcal{N}_{-i} – We say that each of the nodes in $\Gamma_{n,i}$ is paired to node i . Thus, for any subset $A \subseteq \mathcal{N}_{-i}$, we require

$$\mathbb{P}[\Gamma_{n,i} = A] = \begin{cases} \binom{n-1}{K}^{-1} & \text{if } |A| = K \\ 0 & \text{otherwise.} \end{cases}$$

The selection of $\Gamma_{n,i}$ is done *uniformly* amongst all subsets of \mathcal{N}_{-i} which are of size exactly K . The rvs $\Gamma_{n,1}, \dots, \Gamma_{n,n}$ are

assumed to be mutually independent so that

$$\mathbb{P}[\Gamma_{n,i} = A_i, i = 1, \dots, n] = \prod_{i=1}^n \mathbb{P}[\Gamma_{n,i} = A_i]$$

for arbitrary A_1, \dots, A_n subsets of $\mathcal{N}_1, \dots, \mathcal{N}_n$, respectively.

Once this *offline* random pairing has been created, we construct the key rings $\Sigma_{n,1}, \dots, \Sigma_{n,n}$, one for each node, as follows: Assumed available is a collection of nK distinct cryptographic keys $\{\omega_{i|\ell}, i = 1, \dots, n; \ell = 1, \dots, K\}$. Fix $i = 1, \dots, n$ and let $\ell_{n,i} : \Gamma_{n,i} \rightarrow \{1, \dots, K\}$ denote a labeling of $\Gamma_{n,i}$. For each node j in $\Gamma_{n,i}$ paired to i , the cryptographic key $\omega_{i|\ell_{n,i}(j)}$ is associated with j . For instance, if the random set $\Gamma_{n,i}$ is realized as $\{j_1, \dots, j_K\}$ with $1 \leq j_1 < \dots < j_K \leq n$, then an obvious labeling consists in $\ell_{n,i}(j_k) = k$ for each $k = 1, \dots, K$ with key $\omega_{i|k}$ associated with node j_k . Of course other labeling are possible, e.g., according to decreasing labels or according to a random permutation. Finally, the pairwise key $\omega_{n,ij}^* = [\text{Id}_i | \text{Id}_j | \omega_{i|\ell_{n,i}(j)}]$ is constructed and inserted in the memory modules of both nodes i and j . The key $\omega_{n,ij}^*$ is assigned *exclusively* to the pair of nodes i and j , hence the terminology pairwise distribution scheme. The key ring $\Sigma_{n,i}$ of node i is the set

$$\Sigma_{n,i} := \{\omega_{n,ij}^*, j \in \Gamma_{n,i}\} \cup \{\omega_{n,ji}^*, i \in \Gamma_{n,j}\}. \quad (2)$$

If two nodes, say i and j , are within communication range of each other, then they can establish a secure link if at least one of the events $i \in \Gamma_{n,j}$ or $j \in \Gamma_{n,i}$ is taking place. Both events can take place, in which case the memory modules of node i and j both contain the distinct keys $\omega_{n,ij}^*$ and $\omega_{n,ji}^*$. Finally, it is plain by construction that this scheme supports node-to-node authentication.

III. THE MODEL

Under full visibility, this pairwise distribution scheme naturally gives rise to the following class of random graphs: With $n = 2, 3, \dots$ and positive integer $K < n$, we say that the distinct nodes i and j are K -adjacent, written $i \sim_K j$, if and only if they have at least one key in common in their key rings, namely

$$i \sim_K j \quad \text{iff} \quad \Sigma_{n,i} \cap \Sigma_{n,j} \neq \emptyset. \quad (3)$$

Let $\mathbb{H}(n; K)$ denote the undirected random graph on the vertex set $\{1, \dots, n\}$ induced by the adjacency notion (3); this corresponds to modelling the pairwise distribution scheme under full visibility. We have

$$\mathbb{P}[i \sim_K j] = \lambda_n(K) \quad (4)$$

where $\lambda_n(K)$ is the link assignment probability in $\mathbb{H}(n; K)$ given by

$$\begin{aligned} \lambda_n(K) &= 1 - \left(1 - \frac{K}{n-1}\right)^2 \\ &= \frac{2K}{n-1} - \left(\frac{K}{n-1}\right)^2. \end{aligned} \quad (5)$$

As mentioned earlier, in this paper we seek to account for the possibility that communication links between nodes may not be available. To study such situations, we assume a communication model that consists of independent channels each of which can be either on or off. Thus, with p in $(0, 1)$, let $\{B_{ij}(p), 1 \leq i < j \leq n\}$ denote i.i.d. $\{0, 1\}$ -valued rvs with success probability p . The channel between nodes i and j is available (resp. up) with probability p and unavailable (resp. down) with the complementary probability $1 - p$.

Distinct nodes i and j are said to be B -adjacent, written $i \sim_B j$, if $B_{ij}(p) = 1$. The notion of B -adjacency defines the standard ER graph $\mathbb{G}(n; p)$ on the vertex set $\{1, \dots, n\}$. Obviously,

$$\mathbb{P}[i \sim_B j] = p.$$

The random graph model studied here is obtained by *intersecting* the random pairwise graph $\mathbb{H}(n; K)$ with the ER graph $\mathbb{G}(n; p)$. More precisely, the distinct nodes i and j are said to be adjacent, written $i \sim j$, if and only they are both K -adjacent and B -adjacent, namely

$$i \sim j \quad \text{iff} \quad \begin{array}{l} \Sigma_{n,i} \cap \Sigma_{n,j} \neq \emptyset \\ \text{and} \\ B_{ij}(p) = 1. \end{array} \quad (6)$$

The resulting *undirected* random graph defined on the vertex set $\{1, \dots, n\}$ through this notion of adjacency is denoted $\mathbb{H} \cap \mathbb{G}(n; K, p)$.

Throughout the collections of rvs $\{\Gamma_{n,1}, \dots, \Gamma_{n,n}\}$ and $\{B_{ij}(p), 1 \leq i < j \leq n\}$ are assumed to be independent, in which case the edge occurrence probability in $\mathbb{H} \cap \mathbb{G}(n; K, p)$ is given by

$$\mathbb{P}[i \sim j] = p \cdot \mathbb{P}[i \sim_K j] = p\lambda_n(K). \quad (7)$$

IV. THE RESULTS

To fix the terminology, we refer to any mapping $K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ as a *scaling* (for random pairwise graphs) provided it satisfies the natural conditions

$$K_n < n, \quad n = 1, 2, \dots \quad (8)$$

Similarly, any mapping $p : \mathbb{N}_0 \rightarrow (0, 1)$ defines a scaling for ER graphs.

To lighten the notation we often group the parameters K and p into the ordered pair $\theta \equiv (K, p)$. Hence, a mapping $\theta : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \times (0, 1)$ defines a scaling for the intersection graph $\mathbb{H} \cap \mathbb{G}(n; \theta)$ provided the condition (8) holds on the first component.

The results will be expressed in terms of the threshold function $\tau : [0, 1] \rightarrow [0, 1]$ defined by

$$\tau(p) = \begin{cases} 1 & \text{if } p = 0 \\ \frac{2}{1 - \frac{\log(1-p)}{p}} & \text{if } 0 < p < 1 \\ 0 & \text{if } p = 1. \end{cases} \quad (9)$$

It is easy to check that this threshold function is continuous on its entire domain of definition; see Figure 3.

A. Absence of isolated nodes

The first result gives a zero-one law for the absence of isolated nodes.

Theorem 4.1: Consider scalings $K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ and $p : \mathbb{N}_0 \rightarrow (0, 1)$ such that

$$p_n \left(2K_n - \frac{K_n^2}{n-1} \right) \sim c \log n, \quad n = 1, 2, \dots \quad (10)$$

for some $c > 0$. If $\lim_{n \rightarrow \infty} p_n = p^*$ for some p^* in $[0, 1]$, then we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P} \left[\begin{array}{c} \mathbb{H} \cap \mathbb{G}(n; \theta_n) \text{ contains} \\ \text{no isolated nodes} \end{array} \right] \\ &= \begin{cases} 0 & \text{if } c < \tau(p^*) \\ 1 & \text{if } c > \tau(p^*). \end{cases} \end{aligned} \quad (11)$$

The condition (10) on the scaling $\mathbb{N}_0 \rightarrow (0, 1) \times \mathbb{N}_0$ will often be used in the equivalent form

$$p_n \left(2K_n - \frac{K_n^2}{n-1} \right) = c_n \log n, \quad n = 1, 2, \dots \quad (12)$$

with the sequence $c : \mathbb{N}_0 \rightarrow \mathbb{R}_+$ satisfying $\lim_{n \rightarrow \infty} c_n = c$.

B. Connectivity

An analog of Theorem 4.1 also holds for the property of graph connectivity.

Theorem 4.2: Consider scalings $K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ and $p : \mathbb{N}_0 \rightarrow (0, 1)$ such that (10) holds for some $c > 0$. If $\lim_{n \rightarrow \infty} p_n = p^*$ for some p^* in $[0, 1]$, then we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P} [\mathbb{H} \cap \mathbb{G}(n; \theta_n) \text{ is connected}] \\ &= \begin{cases} 0 & \text{if } c < \tau(p^*) \\ 1 & \text{if } c > \tau(p^*) \end{cases} \end{aligned} \quad (13)$$

where the threshold $\tau(p^*)$ is given by (9).

Comparing Theorem 4.2 with Theorem 4.1, we see that the class of random graphs studied here provides one more instance where the zero-one laws for absence of isolated nodes and connectivity coincide, viz. ER graphs [5], random geometric graphs [19] or random key graphs [3], [18], [24].

A case of particular interest arises when $p^* > 0$ since requiring (10) now amounts to

$$\left(2K_n - \frac{K_n^2}{n-1} \right) \sim \frac{c}{p^*} \log n \quad (14)$$

for some $c > 0$. Any scaling $K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ which behaves like (14) must necessarily satisfy $K_n = o(n)$, and it is easy to see that requiring (10) is equivalent to

$$K_n \sim t \log n \quad (15)$$

for some $t > 0$ with c and t related by $t = \frac{c}{2p^*}$. With this reparametrization, Theorem 4.1 and Theorem 4.2 can be summarized in the following simpler form:

Theorem 4.3: Consider scalings $K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ and $p : \mathbb{N}_0 \rightarrow (0, 1)$ such that $\lim_{n \rightarrow \infty} p_n = p^* > 0$. Under the condition (15) for some $t > 0$, we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P} [\mathbb{H} \cap \mathbb{G}(n; \theta_n) \text{ contains no isolated nodes}] \\ &= \lim_{n \rightarrow \infty} \mathbb{P} [\mathbb{H} \cap \mathbb{G}(n; \theta_n) \text{ is connected}] \\ &= \begin{cases} 0 & \text{if } t < \hat{\tau}(p^*) \\ 1 & \text{if } t > \hat{\tau}(p^*) \end{cases} \end{aligned} \quad (16)$$

where we have set

$$\hat{\tau}(p) := \frac{\tau(p)}{2p} = \frac{1}{p - \log(1-p)}, \quad 0 < p < 1. \quad (17)$$

This alternate formulation is particularly relevant for the case $p_n = p^*$ (in $(0, 1)$) for all $n = 1, 2, \dots$, which captures situations when channel conditions are not affected by the number of users. Such simplifications do not occur in the more realistic case $p^* = 0$ which corresponds to the situation where channel conditions are indeed influenced by the number of users in the system – The more users in the network, the more likely they will experience interferences from other users.

We now present numerical results that verify (16). In all the simulations, we fix the number of nodes at $n = 200$. We consider the channel parameters $p = 0.2$, $p = 0.4$, $p = 0.6$, $p = 0.8$, and $p = 1$ (the full visibility case), while varying the parameter K from 1 to 25. For each parameter pair (K, p) , we generate 500 independent samples of the graph $\mathbb{H} \cap \mathbb{G}(n; K, p)$ and count the number of times (out of a possible 500) that the obtained graphs i) have no isolated nodes and ii) are connected. Dividing the counts by 500, we obtain the (empirical) probabilities for the events of interest. The results for connectivity are depicted in Figure 1, where the curve fitting tool of MATLAB is used. It is easy to check that for each value of $p \neq 1$, the connectivity threshold matches the prescription (16), namely $K = \hat{\tau}(p) \log n$. It is also seen that, if the channel is poor, i.e., if p is close to zero, then the required value for K to ensure connectivity can be much larger than the one in the full visibility case $p = 1$. The results regarding the absence of node isolation are depicted in Figure 2. For each value of $p \neq 1$, Figure 2 is indistinguishable from Figure 1, with the difference between the estimated probabilities of graph connectivity and absence of isolated nodes being quite small, in agreement with (16).

V. DISCUSSION AND COMMENTS

A. Comparing with the full-visibility case

At this point the reader may wonder as to what form would Theorem 4.2 take in the context of full visibility – In the setting developed here this corresponds to $p = 1$ so that $\mathbb{H} \cap \mathbb{G}(n; \theta)$ coincides with $\mathbb{H}(n; K)$; see the curve for $p = 1$ in Figure 1). Relevant results for this case were obtained recently by the authors in [25].

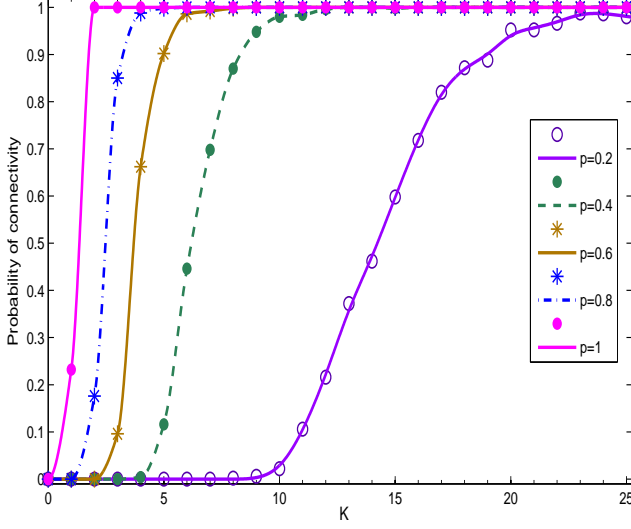


Fig. 1. Probability that $\mathbb{H} \cap \mathbb{G}(n; K, p)$ is connected as a function of K for $p = 0.2, p = 0.4, p = 0.6, p = 0.8$ and $p = 1$ with $n = 200$.

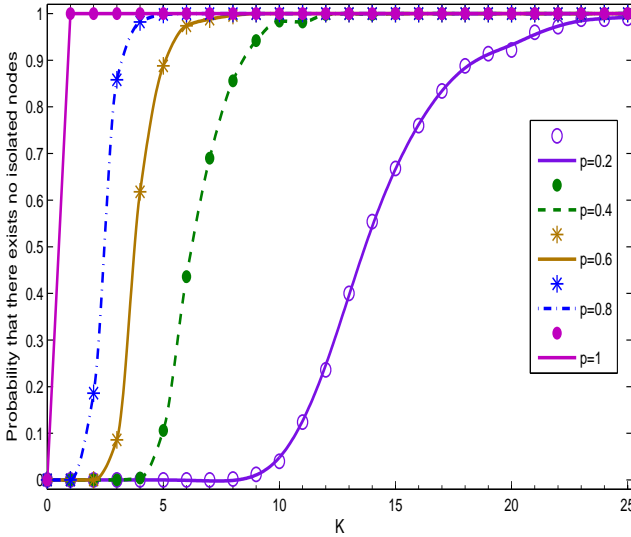


Fig. 2. Probability that $\mathbb{H} \cap \mathbb{G}(n; K, p)$ has no isolated nodes as a function of K for $p = 0.2, p = 0.4, p = 0.6, p = 0.8$ and $p = 1$ with $n = 200$. This figure clearly resembles Figure 1 for all $p \neq 1$.

Theorem 5.1: For any K a positive integer, it holds that

$$\lim_{n \rightarrow \infty} \mathbb{P}[\mathbb{H}(n; K) \text{ is connected}] = \begin{cases} 0 & \text{if } K = 1 \\ 1 & \text{if } K \geq 2. \end{cases}$$

The case where the parameter K is scaled with n is an easy corollary of Theorem 5.1.

Corollary 5.2: For any scaling $K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ such that $K_n \geq 2$ for all n sufficiently large, we have the one-law

$$\lim_{n \rightarrow \infty} \mathbb{P}[\mathbb{H}(n; K_n) \text{ is connected}] = 1.$$

Each node in $\mathbb{H}(n; K)$ has degree at least K , so that no node is ever isolated in $\mathbb{H}(n; K)$. This is in sharp contrast with the model studied here, as reflected by the full zero-one law for node isolation given in Theorem 4.1.

Theorem 5.1 and its Corollary 5.2 together show that very small values of K suffice to ensure asymptotically almost sure (a.a.s.) connectivity of the random graph $\mathbb{H}(n; K)$. However, these two results cannot be recovered from Theorem 4.2 whose zero-one laws are derived under the assumption $p_n < 1$ for all $n = 1, 2, \dots$. Furthermore, even if the scaling $p : \mathbb{N}_0 \rightarrow (0, 1)$ were to satisfy $\lim_{n \rightarrow \infty} p_n = 1$, only the one-laws in Theorem 4.3 remain since $\tau(p^*) = 0$ (and $\hat{\tau}(p^*) = 0$) at $p^* = 1$. Although this might perhaps be expected given the aforementioned absence of isolated nodes in $\mathbb{H}(n; K)$, the one-laws for both the absence of isolated nodes and graph connectivity in $\mathbb{H} \cap \mathbb{G}(n; \theta)$ still require conditions on the behavior of the scaling $K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$, namely (15) (whereas Corollary 5.2 does not).

B. Comparing $\mathbb{H} \cap \mathbb{G}(n; \theta)$ with ER graphs

In the original paper of Chan et al. [7] (as in the reference [13]), the connectivity analysis of the pairwise scheme was based on ER graphs [5] – It was assumed that the random graph induced by the pairwise scheme under a communication model (taken mostly to be the disk model [12]) behaves *like* an ER graph; similar assumptions have been made in [11], [13] when discussing the connectivity of the EG scheme. However, this assumption was made without any formal justification. Recently we have shown that the full visibility model $\mathbb{H}(n; K)$ has major differences with an ER graph. For instance, the edge assignments are (negatively) correlated in $\mathbb{H}(n; K)$ while independent in ER graphs; see [25] for a detailed discussion on the differences of $\mathbb{H}(n; K)$ and $\mathbb{G}(n; p)$. It is easy to verify that the edge assignments in $\mathbb{H} \cap \mathbb{G}(n; \theta)$ are also negatively correlated; see Section IX. Therefore, the models $\mathbb{H}(n; K)$ and $\mathbb{H} \cap \mathbb{G}(n; \theta)$ cannot be equated with an ER graph, and the results obtained in [25] and in this paper are *not* mere consequences of classical results for ER graphs.

However, *formal* similarities do exist between $\mathbb{H} \cap \mathbb{G}(n; \theta)$ and ER graphs. Recall the following well-known zero-one law for ER graphs: For any scaling $p : \mathbb{N}_0 \rightarrow [0, 1]$ satisfying

$$p_n \sim c \frac{\log n}{n}$$

for some $c > 0$, it holds that

$$\lim_{n \rightarrow \infty} \mathbb{P}[\mathbb{G}(n; p_n) \text{ is connected}] = \begin{cases} 0 & \text{if } c < 1 \\ 1 & \text{if } c > 1. \end{cases}$$

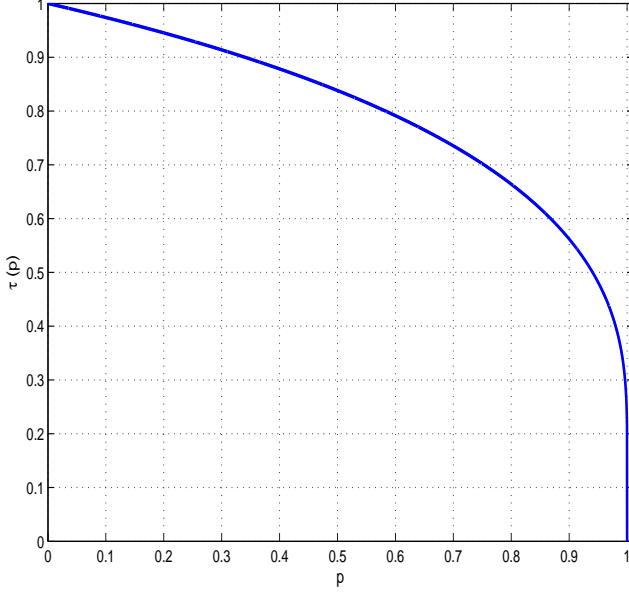


Fig. 3. $\tau(p)$ vs p . Clearly $\tau(p^*) = 1$ only if $\lim_{n \rightarrow \infty} p_n = p^* = 0$.

On the other hand, the condition (10) can be rephrased more compactly as

$$p_n \lambda_n(K_n) \sim c \frac{\log n}{n}, \quad c > 0$$

with the results (11) and (13) unchanged. Hence, in both ER graphs and $\mathbb{H} \cap \mathbb{G}(n; \theta)$, the zero-one laws can be expressed as a comparison of the probability of link assignment against the critical scaling $\frac{\log n}{n}$; this is also the case for random geometric graphs [19], and random key graphs [3], [18], [24]. But the condition $c > \tau(p^*)$ that ensures a.a.s. connectivity in $\mathbb{H} \cap \mathbb{G}(n; \theta)$ is not the same as the condition $c > 1$ for a.a.s. connectivity in ER graphs; see Figure 3. Thus, the connectivity behavior of the model $\mathbb{H} \cap \mathbb{G}(n; \theta)$ is in general different from that in an ER graph, and a “transfer” of the connectivity results from ER graphs cannot be taken for granted. Yet, the comparison becomes intricate when the channel is poor: The connectivity behaviors of the two models do match in the practically relevant case (for WSNs) $\lim_{n \rightarrow \infty} p_n = 0$ since $\tau(0) = 1$.

C. A more realistic communication model

One possible extension of the work presented here would be to consider a more realistic communication model; e.g., the popular disk model [12] which takes into account the geographical positions of the sensor nodes. For instance, assume that the nodes are distributed over a bounded region \mathcal{D} of the plane. According to the *disk model*, nodes i and j located at \mathbf{x}_i and \mathbf{x}_j , respectively, in \mathcal{D} are able to communicate if

$$\|\mathbf{x}_i - \mathbf{x}_j\| < \rho \quad (18)$$

where $\rho > 0$ is called the transmission range. When the node locations are independently and randomly distributed over the

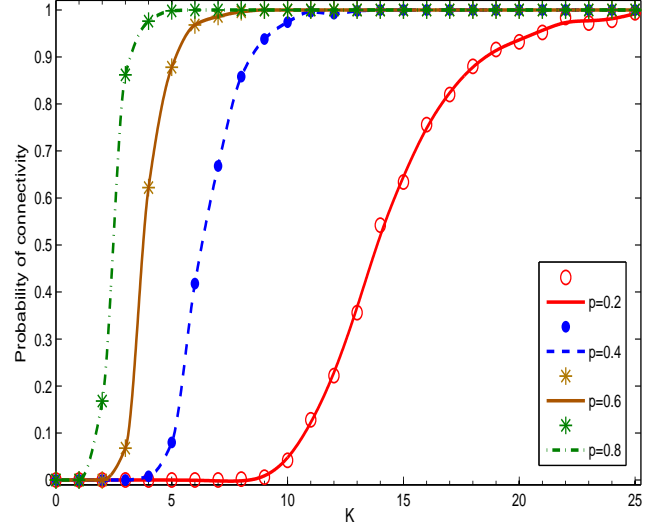


Fig. 4. Probability that $\mathbb{H} \cap \mathbb{G}(n; K, \rho)$ is connected as a function of K . The number of nodes is set to $n = 200$ and ρ is given by $\pi\rho^2 = p$.

region \mathcal{D} , the graph induced under the condition (18) is known as a random geometric graph [19], thereafter denoted $\mathbb{G}(n; \rho)$.

Under the disk model, studying the pairwise scheme of Chan et al. amounts to analyzing the intersection of $\mathbb{H}(n; K)$ and $\mathbb{G}(n; \rho)$, say $\mathbb{H} \cap \mathbb{G}(n; K, \rho)$. A direct analysis of this model seems to be very challenging; see below for more on this. However, limited simulations already suggest that the zero-one laws obtained here for $\mathbb{H} \cap \mathbb{G}(n; K, p)$ have an analog for the model $\mathbb{H} \cap \mathbb{G}(n; K, \rho)$. To verify this, consider 200 nodes distributed uniformly and independently over a folded unit square $[0, 1]^2$ with toroidal (continuous) boundary conditions. Since there are no border effects, it is easy to check that

$$\mathbb{P}[\|\mathbf{x}_i - \mathbf{x}_j\| < \rho] = \pi\rho^2, \quad i \neq j, \quad i, j = 1, 2, \dots, n.$$

whenever $\rho < 0.5$. We match the two communication models $\mathbb{G}(n; p)$ and $\mathbb{G}(n; \rho)$ by requiring $\pi\rho^2 = p$. Then, we use the same procedure that produced Figure 1 to obtain the empirical probability that $\mathbb{H} \cap \mathbb{G}(n; K, \rho)$ is connected for various values of K and p . The results are depicted in Figure 4 whose resemblance with Figure 1 suggests that the connectivity behaviors of the models $\mathbb{H} \cap \mathbb{G}(n; K, p)$ and $\mathbb{H} \cap \mathbb{G}(n; K, \rho)$ are quite similar. This raises the possibility that the results obtained here for the on/off communication model can also be used for dimensioning the pairwise scheme under the disk model.

A complete analysis of $\mathbb{H} \cap \mathbb{G}(n; K, \rho)$ is likely to be very challenging given the difficulties already encountered in the analysis of similar problems. For example, the intersection of random geometric graphs with ER graphs was considered in [2], [28]. Although zero-one laws for graph connectivity are

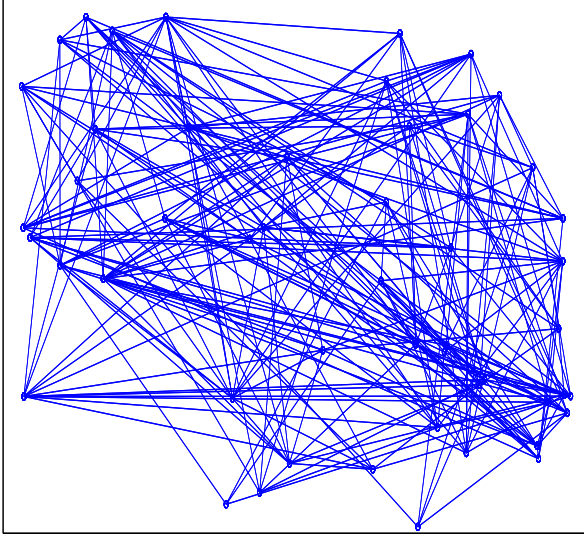


Fig. 5. An instantiation of ER graph $\mathbb{G}(n; p)$ with $n = 50$ and $p = 0.2$.— The graph is connected.

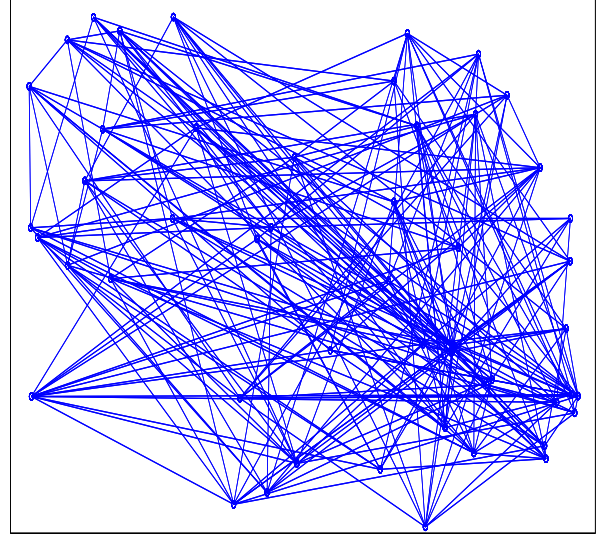


Fig. 6. An instantiation of $\mathbb{H}(n; K)$ with $n = 50$ and $K = 5$.— The graph is connected.

available for each component random graph, the results for the intersection model in [2], [28] were limited only to the absence of isolated nodes; the connectivity problem is still open for that model. Yi et al. [28] also consider the intersection of random key graphs with random geometric graphs, but these results are again limited to the property of node isolation. To the best of our knowledge, Theorem 4.2 reported here constitutes the only zero-one law for graph connectivity in a model formed by intersecting multiple random graphs! (Except of course the trivial case where an ER graph intersects another ER graph.)

D. Intersection of random graphs

When using random graph models to study networks, situations arise where the notion of adjacency between nodes reflects multiple constraints. This can be so even when dealing with networks other than WSNs. As was the case here, such circumstances call for studying models which are constructed by taking the intersection of multiple random graphs. However, as pointed out earlier, the availability of results for each component model does not necessarily imply the availability of results for the intersection of these models; see the examples provided in the previous section.

Figures 5-7 can help better understand the relevant issues as to why this is so: Figure 5 provides a sample of an ER graph $\mathbb{G}(n, p)$ with $n = 200$ and $p = 0.2$. As would be expected from the classical results, the obtained graph is very densely connected. Similarly, Figure 6 provides a sample of the pairwise random graph $\mathbb{H}(n; K)$ with $n = 200$ and $K = 5$. In line with Theorem 5.1, the obtained graph is connected. On the other hand, the graph formed by intersecting these graphs turn out to be *disconnected* as shown in Figure 7.

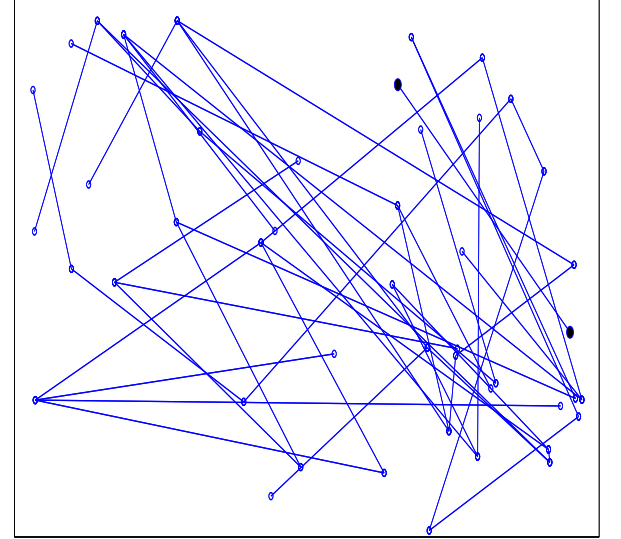


Fig. 7. The intersection $\mathbb{H} \cap \mathbb{G}(n; \theta)$ of the graphs in Figure 5 and Figure 6 – The graph is disconnected as the marked nodes form a component!

To drive this point further, consider the constant parameter case for the models $\mathbb{H}(n; K)$ and $\mathbb{G}(n; p)$, a case which cannot be recovered from either Theorem 4.1 or Theorem 4.2. Nevertheless, Theorem 5.1 yields

$$\lim_{n \rightarrow \infty} \mathbb{P}[\mathbb{H}(n; K) \text{ is connected}] = 1, \quad K \geq 2$$

while it well known [5] that

$$\lim_{n \rightarrow \infty} \mathbb{P}[\mathbb{G}(n; p) \text{ is connected}] = 1. \quad 0 < p < 1.$$

However, it can be shown that

$$\lim_{n \rightarrow \infty} \mathbb{P}[\mathbb{H} \cap \mathbb{G}(n; \theta) \text{ contains no isolated nodes}] = 0 \quad (19)$$

whence

$$\lim_{n \rightarrow \infty} \mathbb{P}[\mathbb{H} \cap \mathbb{G}(n; \theta) \text{ is connected}] = 0 \quad (20)$$

for the same ranges of values for p and K ; for details see the discussion at the end of Section X. This clearly provides a *non-trivial* example (one that is not for an ER intersecting an ER graph) where the intersection of two random graphs is indeed a.s. *not* connected although each of the components is a.s. connected.

VI. A PROOF OF THEOREM 4.1

We prove Theorem 4.1 by the method of first and second moments [14, p. 55] applied to the total number of isolated nodes in $\mathbb{H} \cap \mathbb{G}(n; \theta)$. First some notation: Fix $n = 2, 3, \dots$ and consider $\theta = (K, p)$ with p in $(0, 1)$ and positive integer K such that $K < n$. With

$$\chi_{n,i}(\theta) := \mathbf{1}[\text{Node } i \text{ is isolated in } \mathbb{H} \cap \mathbb{G}(n; \theta)]$$

for each $i = 1, \dots, n$, the number of isolated nodes in $\mathbb{H} \cap \mathbb{G}(n; \theta)$ is simply given by

$$I_n(\theta) := \sum_{i=1}^n \chi_{n,i}(\theta).$$

The random graph $\mathbb{H} \cap \mathbb{G}(n; \theta)$ has no isolated nodes if and only if $I_n(\theta) = 0$.

The method of first moment [14, Eqn (3.10), p. 55] relies on the well-known bound

$$1 - \mathbb{E}[I_n(\theta)] \leq \mathbb{P}[I_n(\theta) = 0] \quad (21)$$

while the method of second moment [14, Remark 3.1, p. 55] has its starting point in the inequality

$$\mathbb{P}[I_n(\theta) = 0] \leq 1 - \frac{\mathbb{E}[I_n(\theta)]^2}{\mathbb{E}[I_n(\theta)]}. \quad (22)$$

The rvs $\chi_{n,1}(\theta), \dots, \chi_{n,n}(\theta)$ being exchangeable, we find

$$\mathbb{E}[I_n(\theta)] = n\mathbb{E}[\chi_{n,1}(\theta)] \quad (23)$$

and

$$\mathbb{E}[I_n(\theta)^2] = n\mathbb{E}[\chi_{n,1}(\theta)] + n(n-1)\mathbb{E}[\chi_{n,1}(\theta)\chi_{n,2}(\theta)]$$

by the binary nature of the rvs involved. It then follows that

$$\begin{aligned} \frac{\mathbb{E}[I_n(\theta)^2]}{\mathbb{E}[I_n(\theta)]^2} &= \frac{1}{n\mathbb{E}[\chi_{n,1}(\theta)]} \\ &\quad + \frac{n-1}{n} \cdot \frac{\mathbb{E}[\chi_{n,1}(\theta)\chi_{n,2}(\theta)]}{(\mathbb{E}[\chi_{n,1}(\theta)])^2}. \end{aligned} \quad (24)$$

From (21) and (23) we see that the one-law $\lim_{n \rightarrow \infty} \mathbb{P}[I_n(\theta_n) = 0] = 1$ will be established if we show that

$$\lim_{n \rightarrow \infty} n\mathbb{E}[\chi_{n,1}(\theta_n)] = 0. \quad (25)$$

It is also plain from (22) and (24) that the zero-law $\lim_{n \rightarrow \infty} \mathbb{P}[I_n(\theta_n) = 0] = 0$ holds if

$$\lim_{n \rightarrow \infty} n\mathbb{E}[\chi_{n,1}(\theta_n)] = \infty \quad (26)$$

and

$$\limsup_{n \rightarrow \infty} \left(\frac{\mathbb{E}[\chi_{n,1}(\theta_n)\chi_{n,2}(\theta_n)]}{(\mathbb{E}[\chi_{n,1}(\theta_n)])^2} \right) \leq 1. \quad (27)$$

The proof of Theorem 4.1 passes through the next two technical propositions which establish (25), (26) and (27) under the appropriate conditions on the scaling $\theta : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \times (0, 1)$.

Proposition 6.1: Consider scalings $K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ and $p : \mathbb{N}_0 \rightarrow (0, 1)$ such that (10) holds for some $c > 0$. Assume also that $\lim_{n \rightarrow \infty} p_n = p^*$ exists. Then, we have

$$\lim_{n \rightarrow \infty} n\mathbb{E}[\chi_{n,1}(\theta_n)] = \begin{cases} 0 & \text{if } c > \tau(p^*) \\ \infty & \text{if } c < \tau(p^*) \end{cases} \quad (28)$$

where the threshold $\tau(p^*)$ is given by (9).

A proof of Proposition 6.1 is given in Section VIII.

Proposition 6.2: Consider scalings $K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ and $p : \mathbb{N}_0 \rightarrow (0, 1)$ such that (10) holds for some $c > 0$. Assume also that $\lim_{n \rightarrow \infty} p_n = p^*$ exists. Then, we have (27) whenever $p^* < 1$.

A proof of Proposition 6.2 can be found in Section X. To complete the proof of Theorem 4.1, pick a scaling $\theta : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \times (0, 1)$ such that (10) holds for some $c > 0$ and $\lim_{n \rightarrow \infty} p_n = p^*$ exists. Under the condition $c > \tau(p^*)$ we get (25) from Proposition 6.1, and the one-law $\lim_{n \rightarrow \infty} \mathbb{P}[I_n(\theta_n) = 0] = 1$ follows. Next, assume that $c < \tau(p^*)$ – This case is possible only if $p^* < 1$ since $\tau(1) = 0$ as seen at (9). When $p^* < 1$, we obtain (26) and (27) with the help of Propositions 6.1 and 6.2, respectively. The conclusion $\lim_{n \rightarrow \infty} \mathbb{P}[I_n(\theta_n) = 0] = 0$ is now immediate.

VII. A PREPARATORY RESULT

Fix $n = 2, 3, \dots$ and consider $\theta = (K, p)$ with p in $(0, 1)$ and positive integer K such that $K < n$. Under the enforced assumptions, for all $i = 1, \dots, n$, we easily see that

$$\mathbb{E}[\chi_{n,i}(\theta)] = \mathbb{E}[(1-p)^{D_{n,i}}] \quad (29)$$

where $D_{n,i}$ denotes the degree of node i in $\mathbb{H}(n; K)$. Note that

$$D_{n,i} = K + \sum_{j=1, j \neq i}^n \mathbf{1}[i \in \Gamma_{n,j}]. \quad (30)$$

By independence, since

$$|\{j = 1, \dots, n : j \notin \Gamma_{n,i} \cup \{i\}\}| = n - K - 1,$$

the second term in (30) is a binomial rv with $n - K - 1$ trials and success probability given by

$$\mathbb{P}[i \in \Gamma_{n,j}] = \frac{\binom{n-2}{K-1}}{\binom{n-1}{K}} = \frac{K}{n-1}, \quad (31)$$

whence

$$\mathbb{E}[\chi_{n,i}(\theta)] = (1-p)^K \cdot \left(1 - \frac{pK}{n-1}\right)^{n-K-1}. \quad (32)$$

The proof of Proposition 6.1 uses a somewhat simpler form of the expression (32) which we develop next.

Lemma 7.1: Consider scalings $K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ and $p : \mathbb{N}_0 \rightarrow (0, 1)$ such that (10) holds for some $c > 0$. It holds that

$$n\mathbb{E}[\chi_{n,1}(\theta_n)] = e^{\alpha_n + o(1)} \quad n = 1, 2, \dots \quad (33)$$

with

$$\alpha_n := (1 - c_n) \log n + K_n(p_n + \log(1 - p_n)) \quad (34)$$

where the sequence $c : \mathbb{N}_0 \rightarrow \mathbb{R}$ is the one appearing in the form (12) of the condition (10).

In what follows we make use of the decomposition

$$\log(1 - x) = -x - \Psi(x), \quad 0 \leq x < 1 \quad (35)$$

with

$$\Psi(x) := \int_0^x \frac{t}{1-t} dt$$

on that range. Note that

$$\lim_{x \downarrow 0} \frac{\Psi(x)}{x^2} = \frac{1}{2}.$$

Proof. Consider a scaling $\theta : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \times (0, 1)$ such that (10) holds for some $c > 0$ and assume the existence of the limit $\lim_{n \rightarrow \infty} p_n = p^*$. Replacing θ by θ_n in (32) for each $n = 2, 3, \dots$ we get

$$n\mathbb{E}[\chi_{n,1}(\theta_n)] = e^{\beta_n} \quad (36)$$

with β_n given by

$$\beta_n = \log n + K_n \log(1 - p_n) - \gamma_n$$

with

$$\gamma_n := -(n - K_n - 1) \log \left(1 - \frac{p_n K_n}{n-1}\right).$$

The decomposition (35) now yields

$$\begin{aligned} \gamma_n &:= (n - K_n - 1) \left(\frac{p_n K_n}{n-1} + \Psi \left(\frac{p_n K_n}{n-1} \right) \right) \\ &= \left(1 - \frac{K_n}{n-1} \right) K_n p_n + (n - K_n - 1) \Psi \left(\frac{p_n K_n}{n-1} \right) \\ &= -K_n p_n + \left(2 - \frac{K_n}{n-1} \right) K_n p_n \\ &\quad + (n - K_n - 1) \Psi \left(\frac{p_n K_n}{n-1} \right) \\ &= -K_n p_n + c_n \log n + (n - K_n - 1) \Psi \left(\frac{p_n K_n}{n-1} \right) \end{aligned}$$

where the last step used the form (12) of the condition (10) on the scaling. Reporting this calculation into the expression for β_n we find

$$\beta_n = \alpha_n - (n - K_n - 1) \Psi \left(\frac{p_n K_n}{n-1} \right).$$

Lemma 7.1 will be established if we show that

$$\lim_{n \rightarrow \infty} (n - K_n - 1) \Psi \left(\frac{p_n K_n}{n-1} \right) = 0. \quad (37)$$

To that end, for each $n = 2, 3, \dots$ we note that

$$p_n K_n \leq p_n \left(2K_n - \frac{K_n^2}{n-1} \right) \leq 2p_n K_n$$

since $K_n < n$. The condition (12) implies

$$\frac{c_n}{2} \log n \leq p_n K_n \leq c_n \log n, \quad (38)$$

and it is now plain that

$$\lim_{n \rightarrow \infty} \frac{p_n K_n}{n-1} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} (n - K_n - 1) \frac{p_n^2 K_n^2}{(n-1)^2} = 0.$$

Invoking the behavior of $\Psi(x)$ at $x = 0$ mentioned earlier, we conclude from these facts that

$$\lim_{n \rightarrow \infty} \left((n - K_n - 1) \frac{p_n^2 K_n^2}{(n-1)^2} \right) \left(\frac{\Psi \left(\frac{p_n K_n}{n-1} \right)}{\left(\frac{p_n K_n}{n-1} \right)^2} \right) = 0. \quad (39)$$

This establishes (37) and the proof of Lemma 7.1 is completed. \blacksquare

VIII. A PROOF OF PROPOSITION 6.1

In view of Lemma 7.1, Proposition 6.1 will be established if we show

$$\lim_{n \rightarrow \infty} \alpha_n = \begin{cases} -\infty & \text{if } c > \tau(p^*) \\ +\infty & \text{if } c < \tau(p^*). \end{cases} \quad (40)$$

To see this, first note from (35) that for each $n = 1, 2, \dots$, we have $p_n + \log(1 - p_n) \leq 0$ and the lower bound in (38) implies

$$\begin{aligned} \alpha_n &\leq (1 - c_n) \log n + c_n \left(\frac{\log n}{2p_n} \right) \cdot (p_n + \log(1 - p_n)) \\ &= \left(1 - \frac{c_n}{2} \left(1 - \frac{\log(1 - p_n)}{p_n} \right) \right) \cdot \log n. \end{aligned} \quad (41)$$

Letting n go to infinity in this last expression, we get $\lim_{n \rightarrow \infty} \alpha_n = -\infty$ whenever

$$c > \lim_{n \rightarrow \infty} \frac{2}{1 - \frac{\log(1 - p_n)}{p_n}} = \tau(p^*) \quad (42)$$

since $\lim_{n \rightarrow \infty} c_n = c$.

Next, we show that if $c < \tau(p^*)$, then $\lim_{n \rightarrow \infty} \alpha_n = +\infty$. We only need to consider the case $0 \leq p^* < 1$ since $\tau(1) = 0$

and the constraint $c < \tau(1)$ is vacuous. We begin by assuming $p^* = 0$, in which case for each $n = 2, 3, \dots$, we have

$$\begin{aligned}
\alpha_n &= (1 - c_n) \log n + K_n(p_n + (-p_n - \Psi(p_n))) \\
&= (1 - c_n) \log n - K_n \Psi(p_n) \\
&= (1 - c_n) \log n - \left(\frac{\Psi(p_n)}{p_n^2} \right) \cdot K_n p_n^2 \\
&\geq (1 - c_n) \log n - c_n \log n \cdot \left(\frac{\Psi(p_n)}{p_n^2} \right) p_n \\
&= \log n \cdot \left(1 - c_n \left(1 + \left(\frac{\Psi(p_n)}{p_n^2} \right) p_n \right) \right) \quad (43)
\end{aligned}$$

with the inequality following from the upper bound in (38). Let n grow large in the last expression. Since we have assumed $\lim_{n \rightarrow \infty} p_n = 0$, we get

$$\lim_{n \rightarrow \infty} p_n \left(\frac{\Psi(p_n)}{p_n^2} \right) = 0,$$

and the desired conclusion $\lim_{n \rightarrow \infty} \alpha_n = +\infty$ is obtained whenever $c < 1 = \tau(0)$ upon using $\lim_{n \rightarrow \infty} c_n = c$.

Finally we assume $0 < p^* < 1$. For each $\varepsilon > 0$, there exists a finite positive integer $n^*(\varepsilon)$ such that $p_n \geq (1 - \varepsilon)p^*$ when $n \geq n^*(\varepsilon)$. On that range the upper bound in (38) yields

$$K_n \leq \frac{c}{(1 - \varepsilon)p^*} \cdot \log n,$$

whence the conclusions $K_n^2 = o(n)$ and

$$p_n \left(2K_n - \frac{K_n^2}{n-1} \right) = 2K_n p_n + o(1)$$

follow. Comparing this last fact against the lefthand side of (12) yields

$$K_n p_n = \frac{c_n}{2} \log n + o(1),$$

so that

$$K_n p_n \sim \frac{c_n}{2} \log n. \quad (44)$$

From (34) it follows that

$$\frac{\alpha_n}{\log n} = (1 - c_n) + \left(1 + \frac{\log(1 - p_n)}{p_n} \right) \cdot \frac{K_n p_n}{\log n}$$

for all n sufficiently large. Letting n go to infinity in this last expression and using (44) with the earlier remarks, we readily conclude

$$\lim_{n \rightarrow \infty} \frac{\alpha_n}{\log n} = (1 - c) + \frac{c}{2} \left(1 + \frac{\log(1 - p^*)}{p^*} \right) = 1 - \frac{c}{\tau(p^*)}$$

where the last step follows by direct inspection. It is now clear that $\lim_{n \rightarrow \infty} \alpha_n = \infty$ when $c < \tau(p^*)$ with $0 < p^* < 1$. The proof of Proposition 6.1 is now completed. ■

IX. NEGATIVE DEPENDENCE AND CONSEQUENCES

Fix positive integers $n = 2, 3, \dots$ and K with $K < n$. Several properties of the $\{0, 1\}$ -valued rvs

$$\left\{ \mathbf{1}[j \in \Gamma_{n,i}], \quad i \neq j, \quad i, j = 1, \dots, n \right\} \quad (45)$$

and

$$\left\{ \mathbf{1}[j \in \Gamma_{n,i} \vee i \in \Gamma_{n,j}], \quad i \neq j, \quad i, j = 1, \dots, n \right\} \quad (46)$$

will play a key role in some of the forthcoming arguments.

A. Negative association

The properties of interest can be couched in terms of *negative association*, a form of negative correlation introduced to Joag-Dev and Proschan [15]. We first develop the needed definitions and properties: Let $\{X_\lambda, \lambda \in \Lambda\}$ be a collection of \mathbb{R} -valued rvs indexed by the finite set Λ . For any non-empty subset A of Λ , we write X_A to denote the $\mathbb{R}^{|A|}$ -valued $X_A = (X_\lambda, \lambda \in A)$. The rvs $\{X_\lambda, \lambda \in \Lambda\}$ are then said to be *negatively associated* if for any non-overlapping subsets A and B of Λ and for any monotone increasing mappings $\varphi : \mathbb{R}^{|A|} \rightarrow \mathbb{R}$ and $\psi : \mathbb{R}^{|B|} \rightarrow \mathbb{R}$, the covariance inequality

$$\mathbb{E}[\varphi(X_A)\psi(X_B)] \leq \mathbb{E}[\varphi(X_A)] \mathbb{E}[\psi(X_B)] \quad (47)$$

holds whenever the expectations in (47) are well defined and finite. Note that φ and ψ need only be monotone increasing on the support of X_A and X_B , respectively.

This definition has some easy consequences to be used repeatedly in what follows: The negative association of $\{X_\lambda, \lambda \in \Lambda\}$ implies the negative association of the collection $\{X_\lambda, \lambda \in \Lambda'\}$ where Λ' is any subset of Λ . It is also well known [15, P2, p. 288] that the negative association of the rvs $\{X_\lambda, \lambda \in \Lambda\}$ implies the inequality

$$\mathbb{E} \left[\prod_{\lambda \in A} f_\lambda(X_\lambda) \right] \leq \prod_{\lambda \in A} \mathbb{E}[f_\lambda(X_\lambda)] \quad (48)$$

where A is a subset of Λ and the collection $\{f_\lambda, \lambda \in A\}$ of mappings $\mathbb{R} \rightarrow \mathbb{R}_+$ are all monotone increasing; by non-negativity all the expectations exist and finiteness is moot.

We can apply these ideas to collections of indicator rvs, namely for each λ in Λ , $X_\lambda = \mathbf{1}[E_\lambda]$ for some event E_λ . From the definitions, it is easy to see that if the rvs $\{\mathbf{1}[E_\lambda], \lambda \in \Lambda\}$ are negatively associated, so are the rvs $\{\mathbf{1}[E_\lambda^c], \lambda \in \Lambda\}$. Moreover, for any subset A of Λ , we have

$$\mathbb{P}[E_\lambda, \lambda \in A] \leq \prod_{\lambda \in A} \mathbb{P}[E_\lambda]. \quad (49)$$

This follows from (48) by taking $f_\lambda(x) = x^+$ on \mathbb{R} for each λ in Λ .

B. Useful consequences

A key observation for our purpose is as follows: For each $i = 1, \dots, n$, the rvs

$$\{\mathbf{1}[j \in \Gamma_{n,i}], j \in \mathcal{N}_{-i}\} \quad (50)$$

form a collection of negatively associated rvs. This is a consequence of the fact that the random set $\Gamma_{n,i}$ represents a random sample (without replacement) of size K from \mathcal{N}_{-i} ; see [15, Example 3.2(c)] for details.

The n collections (50) are mutually independent, so that by the ‘‘closure under products’’ property of negative association [15, P7, p. 288] [10, p. 35], the rvs (45) also form a collection of negatively associated rvs.

Hence, by taking complements, the rvs

$$\left\{ \mathbf{1}[j \notin \Gamma_{n,i}], \quad i \neq j, \quad i, j = 1, \dots, n \right\} \quad (51)$$

also form a collection of negatively associated rvs. With distinct $i, j = 1, \dots, n$, we note that

$$\mathbf{1}[i \notin \Gamma_{n,j}, j \notin \Gamma_{n,i}] = f(\mathbf{1}[i \notin \Gamma_{n,j}], \mathbf{1}[j \notin \Gamma_{n,i}]) \quad (52)$$

with mapping $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ given by $f(x, y) = x^+ y^+$ for all x, y in \mathbb{R} . This mapping being non-decreasing on \mathbb{R}^2 , it follows [15, P6, p. 288] that the rvs

$$\left\{ \mathbf{1}[j \notin \Gamma_{n,i}, i \notin \Gamma_{n,j}], \quad i \neq j, \quad i, j = 1, \dots, n \right\} \quad (53)$$

are also negatively associated. Taking complements one more time, we see that the rvs (46) are also negatively associated.

For each $k = 1, 2$ and $j = 3, \dots, n$, we shall find it useful to define

$$u_{n,j,k}(\theta) := \mathbb{E} \left[(1-p)^{\mathbf{1}[k \in \Gamma_{n,j}]} \right]$$

and

$$b_{n,j}(\theta) := \mathbb{E} \left[(1-p)^{\mathbf{1}[1 \in \Gamma_{n,j}] + \mathbf{1}[2 \in \Gamma_{n,j}]} \right].$$

Under the enforced assumptions, we have $b_{n,3}(\theta) = \dots = b_{n,n}(\theta) \equiv b_n(\theta)$ and $u_{n,3,1}(\theta) = \dots = u_{n,n,1}(\theta) = u_{n,3,2}(\theta) = \dots = u_{n,n,2}(\theta) \equiv u_n(\theta)$.

Before computing either one of the quantities $u_n(\theta)$ and $b_n(\theta)$, we note that

$$b_n(\theta) \leq u_n(\theta)^2. \quad (54)$$

This is a straightforward consequence of the negative association of the rvs (45) – In (47), with A and B singletons, use the increasing functions $\varphi, \psi : \mathbb{R} \rightarrow \mathbb{R} : x \rightarrow -(1-p)^x$.

Using (31) we get

$$\begin{aligned} u_n(\theta) &= (1-p) \frac{K}{n-1} + \left(1 - \frac{K}{n-1} \right) \\ &= 1 - p \frac{K}{n-1}. \end{aligned} \quad (55)$$

An expression for $b_n(\theta)$ is available but will not be needed due to the availability of (54).

X. A PROOF OF PROPOSITION 6.2

As expected, the first step in proving Proposition 6.2 consists in evaluating the cross moment appearing in the numerator of (27). Fix $n = 2, 3, \dots$ and consider $\theta = (K, p)$ with p in $(0, 1)$ and positive integer K such that $K < n$. Define the \mathbb{N}_0 -valued rvs $B_n(\theta)$ and $U_n(\theta)$ by

$$B_n(\theta) := \sum_{j=3}^n \mathbf{1}[j \notin \Gamma_{n,1}] \mathbf{1}[j \notin \Gamma_{n,2}] \quad (56)$$

and

$$\begin{aligned} U_n(\theta) &:= \sum_{j=3}^n \mathbf{1}[j \notin \Gamma_{n,1}] \mathbf{1}[j \in \Gamma_{n,2}] \\ &\quad + \sum_{j=3}^n \mathbf{1}[j \notin \Gamma_{n,2}] \mathbf{1}[j \in \Gamma_{n,1}]. \end{aligned} \quad (57)$$

Proposition 10.1: Fix $n = 2, 3, \dots$. For any p in $(0, 1)$ and positive integer K such that $K < n$, we have

$$\begin{aligned} \mathbb{E}[\chi_{n,1}(\theta) \chi_{n,2}(\theta)] &= (1-p)^{2K} \mathbb{E} \left[\frac{b_n(\theta)^{B_n(\theta)} \cdot u_n(\theta)^{U_n(\theta)}}{(1-p)^{\mathbf{1}[2 \in \Gamma_{n,1}, 1 \in \Gamma_{n,2}]}} \right] \end{aligned} \quad (58)$$

where the rvs $B_n(\theta)$ and $U_n(\theta)$ given by (56) and (57), respectively.

A proof of Proposition 10.1 is available in Appendix A. Still in the setting of Proposition 10.1, we can use (54) in conjunction with (58) to get

$$\begin{aligned} \mathbb{E}[\chi_{n,1}(\theta) \chi_{n,2}(\theta)] &\leq (1-p)^{2K} \mathbb{E} \left[\frac{u_n(\theta)^{2B_n(\theta) + U_n(\theta)}}{(1-p)^{\mathbf{1}[2 \in \Gamma_{n,1}, 1 \in \Gamma_{n,2}]} \right]. \end{aligned} \quad (59)$$

It is plain that

$$\begin{aligned} 2B_n(\theta) + U_n(\theta) &= \sum_{j=3}^n \mathbf{1}[j \notin \Gamma_{n,1}] + \sum_{j=3}^n \mathbf{1}[j \notin \Gamma_{n,2}]. \end{aligned}$$

We note that

$$\begin{aligned} \sum_{j=3}^n \mathbf{1}[j \notin \Gamma_{n,1}] &= \sum_{j=2}^n \mathbf{1}[j \notin \Gamma_{n,1}] - \mathbf{1}[2 \notin \Gamma_{n,1}] \\ &= (n-1-K) - (1 - \mathbf{1}[2 \in \Gamma_{n,1}]) \\ &= (n-2-K) + \mathbf{1}[2 \in \Gamma_{n,1}] \end{aligned}$$

and

$$\sum_{j=3}^n \mathbf{1}[j \notin \Gamma_{n,2}] = (n-2-K) + \mathbf{1}[1 \in \Gamma_{n,2}]$$

by similar arguments. The expression

$$\begin{aligned} 2B_n(\theta) + U_n(\theta) &= 2(n-2-K) + \mathbf{1}[2 \in \Gamma_{n,1}] + \mathbf{1}[1 \in \Gamma_{n,2}] \end{aligned} \quad (60)$$

now follows, and we find

$$\begin{aligned} \mathbb{E}[\chi_{n,1}(\theta) \chi_{n,2}(\theta)] &\leq (1-p)^{2K} u_n(\theta)^{2(n-2-K)} \cdot R_n(\theta) \end{aligned} \quad (61)$$

with

$$R_n(\theta) := \mathbb{E} \left[\frac{u_n(\theta)^{\mathbf{1}[2 \in \Gamma_{n,1}] + \mathbf{1}[1 \in \Gamma_{n,2}]} }{(1-p)^{\mathbf{1}[2 \in \Gamma_{n,1}, 1 \in \Gamma_{n,2}]} } \right].$$

Next, with the help of (32) and (55) we conclude that

$$\begin{aligned} & \frac{\mathbb{E} [\chi_{n,1}(\theta) \chi_{n,2}(\theta)]}{(\mathbb{E} [\chi_{n,1}(\theta)])^2} \\ & \leq \frac{(1-p)^{2K} \cdot u_n(\theta)^{2(n-2-K)}}{((1-p)^K \cdot u_n(\theta)^{n-1-K})^2} \cdot R_n(\theta) \\ & = u_n(\theta)^{-2} R_n(\theta) \\ & = \mathbb{E} \left[\frac{u_n(\theta)^{\mathbf{1}[2 \in \Gamma_{n,1}] + \mathbf{1}[1 \in \Gamma_{n,2}] - 2}}{(1-p)^{\mathbf{1}[2 \in \Gamma_{n,1}, 1 \in \Gamma_{n,2}]} } \right]. \end{aligned} \quad (62)$$

Direct inspection readily yields

$$\begin{aligned} & \frac{u_n(\theta)^{\mathbf{1}[2 \in \Gamma_{n,1}] + \mathbf{1}[1 \in \Gamma_{n,2}] - 2}}{(1-p)^{\mathbf{1}[2 \in \Gamma_{n,1}, 1 \in \Gamma_{n,2}]} } \\ & = \begin{cases} \frac{1}{1-p} & \text{if } 2 \in \Gamma_{n,1}, 1 \in \Gamma_{n,2} \\ \left(1 - \frac{pK}{n-1}\right)^{-2} & \text{if } 2 \notin \Gamma_{n,1}, 1 \notin \Gamma_{n,2} \\ \left(1 - \frac{pK}{n-1}\right)^{-1} & \text{otherwise.} \end{cases} \end{aligned} \quad (63)$$

Taking expectation and reporting into (62) we then find

$$\begin{aligned} & \frac{\mathbb{E} [\chi_{n,1}(\theta) \chi_{n,2}(\theta)]}{(\mathbb{E} [\chi_{n,1}(\theta)])^2} \\ & \leq \frac{1}{1-p} \mathbb{P}[2 \in \Gamma_{n,1}, 1 \in \Gamma_{n,2}] + \left(1 - p \frac{K}{n-1}\right)^{-2} \\ & = \frac{1}{1-p} \left(\frac{K}{n-1}\right)^2 + \left(1 - p \frac{K}{n-1}\right)^{-2} \end{aligned} \quad (64)$$

by a crude bounding argument.

Now consider a scaling $\theta : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \times (0, 1)$ such that (10) holds for some $c > 0$ and $\lim_{n \rightarrow \infty} p_n = p^* < 1$. Replace θ by θ_n in the bound (65) with respect to this scaling. It is immediate that (27) will be established if we show that

$$\lim_{n \rightarrow \infty} \frac{1}{1-p_n} \left(\frac{K_n}{n-1}\right)^2 = 0$$

and that

$$\lim_{n \rightarrow \infty} \left(1 - p_n \frac{K_n}{n-1}\right) = 1.$$

These limits are an easy consequence of the inequalities (38) by virtue of the fact that $\lim_{n \rightarrow \infty} p_n = p^* < 1$. ■

We close with a proof of (19): Consider $\theta = (K, p)$ with p in $(0, 1)$ and positive integer K . It follows from (32) that

$$\lim_{n \rightarrow \infty} \mathbb{E} [\chi_{n,1}(\theta)] = (1-p)^K e^{-pK},$$

whence $\lim_{n \rightarrow \infty} \mathbb{E} [I_n(\theta)] = \infty$. It also immediate from (65) that

$$\limsup_{n \rightarrow \infty} \frac{\mathbb{E} [\chi_{n,1}(\theta) \chi_{n,2}(\theta)]}{(\mathbb{E} [\chi_{n,1}(\theta)])^2} \leq 1.$$

The arguments outlined in Section VI now yield

$$\lim_{n \rightarrow \infty} \mathbb{P} [I_n(\theta) = 0] = 0,$$

and this establishes (19). The conclusion (20) immediately follows; see discussion at (66).

XI. A PROOF OF THEOREM 4.2 (PART I)

Fix $n = 2, 3, \dots$ and consider $\theta = (K, p)$ with p in $(0, 1)$ and positive integer K such that $K < n$. We define the events

$$C_n(\theta) := [\mathbb{H} \cap \mathbb{G}(n; \theta) \text{ is connected}]$$

and

$$I(n; \theta) := [\mathbb{H} \cap \mathbb{G}(n; \theta) \text{ contains no isolated nodes}].$$

If the random graph $\mathbb{H} \cap \mathbb{G}(n; \theta)$ is connected, then it does not contain any isolated node, whence $C_n(\theta)$ is a subset of $I(n; \theta)$, and the conclusions

$$\mathbb{P} [C_n(\theta)] \leq \mathbb{P} [I(n; \theta)] \quad (66)$$

and

$$\mathbb{P} [C_n(\theta)^c] = \mathbb{P} [C_n(\theta)^c \cap I(n; \theta)] + \mathbb{P} [I(n; \theta)^c] \quad (67)$$

obtain.

Taken together with Theorem 4.1, the relations (66) and (67) pave the way to proving Theorem 4.2. Indeed, pick a scaling $\theta : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \times (0, 1)$ such that (10) holds for some $c > 0$ and $\lim_{n \rightarrow \infty} p_n = p^*$ exists. If $c < \tau(p^*)$, then $\lim_{n \rightarrow \infty} \mathbb{P} [I(n; \theta_n)] = 0$ by the zero-law for the absence of isolated nodes, whence $\lim_{n \rightarrow \infty} \mathbb{P} [C_n(\theta_n)] = 0$ with the help of (66). If $c > \tau(p^*)$, then $\lim_{n \rightarrow \infty} \mathbb{P} [I(n; \theta_n)] = 1$ by the one-law for the absence of isolated nodes, and the desired conclusion $\lim_{n \rightarrow \infty} \mathbb{P} [C_n(\theta_n)] = 1$ (or equivalently, $\lim_{n \rightarrow \infty} \mathbb{P} [C_n(\theta_n)^c] = 0$) will follow via (67) if we show the following:

Proposition 11.1: For any scaling $\theta : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \times (0, 1)$ such that $\lim_{n \rightarrow \infty} p_n = p^$ exists and (10) holds for some $c > \tau(p^*)$, we have*

$$\lim_{n \rightarrow \infty} \mathbb{P} [C_n(\theta_n)^c \cap I(n; \theta_n)] = 0. \quad (68)$$

The proof of Proposition 11.1 starts below and runs through two more sections, namely Sections XII and XIII. The basic idea is to find a sufficiently tight upper bound on the probability in (68) and then to show that this bound goes to zero as n becomes large. This approach is similar to the one used for proving the one-law for connectivity in ER graphs [5, p. 164].

We begin by finding the needed upper bound: Fix $n = 2, 3, \dots$ and consider $\theta = (K, p)$ with p in $(0, 1)$ and positive integer K such that $K < n$. For any non-empty subset S of nodes, i.e., $S \subseteq \{1, \dots, n\}$, we define the graph $\mathbb{H} \cap \mathbb{G}(n; \theta)(S)$ (with vertex set S) as the subgraph of $\mathbb{H} \cap \mathbb{G}(n; \theta)$ restricted to the nodes in S . We also say that S is *isolated* in $\mathbb{H} \cap \mathbb{G}(n; \theta)$ if there are no edges (in $\mathbb{H} \cap \mathbb{G}(n; \theta)$)

between the nodes in S and the nodes in the complement $S^c = \{1, \dots, n\} - S$. This is characterized by

$$\Sigma_{n,i} \cap \Sigma_{n,j} = \emptyset \quad \forall \quad B_{ij}(p) = 0, \quad i \in S, j \in S^c.$$

With each non-empty subset S of nodes, we associate several events of interest: Let $C_n(\theta; S)$ denote the event that the subgraph $\mathbb{H} \cap \mathbb{G}(n; \theta)(S)$ is itself connected. The event $C_n(\theta; S)$ is completely determined by the rvs $\{K_i(\theta), i \in S\}$. We also introduce the event $B_n(\theta; S)$ to capture the fact that S is isolated in $\mathbb{H} \cap \mathbb{G}(n; \theta)$, i.e.,

$$\begin{aligned} B_n(\theta; S) \\ := [\Sigma_{n,i} \cap \Sigma_{n,j} = \emptyset \quad \forall \quad B_{ij}(p) = 0, \quad i \in S, j \in S^c]. \end{aligned}$$

Finally, we set

$$A_n(\theta; S) := C_n(\theta; S) \cap B_n(\theta; S).$$

The starting point of the discussion is the following basic observation: If $\mathbb{H} \cap \mathbb{G}(n; \theta)$ is *not* connected and yet has *no* isolated nodes, then there must exist a subset S of nodes with $|S| \geq 2$ such that $\mathbb{H} \cap \mathbb{G}(n; \theta)(S)$ is connected while S is isolated in $\mathbb{H} \cap \mathbb{G}(n; \theta)$. This is captured by the inclusion

$$C_n(\theta)^c \cap I(n; \theta) \subseteq \bigcup_{S \subseteq \mathcal{N}: |S| \geq 2} A_n(\theta; S) \quad (69)$$

A moment of reflection should convince the reader that this union need only be taken over all subsets S of $\{1, \dots, n\}$ with $2 \leq |S| \leq \lfloor \frac{n}{2} \rfloor$. A standard union bound argument immediately gives

$$\begin{aligned} \mathbb{P}[C_n(\theta)^c \cap I(n; \theta)] &\leq \sum_{S \subseteq \mathcal{N}: 2 \leq |S| \leq \lfloor \frac{n}{2} \rfloor} \mathbb{P}[A_n(\theta; S)] \\ &= \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \left(\sum_{S \in \mathcal{N}_{n,r}} \mathbb{P}[A_n(\theta; S)] \right) \quad (70) \end{aligned}$$

where $\mathcal{N}_{n,r}$ denotes the collection of all subsets of $\{1, \dots, n\}$ with exactly r elements.

For each $r = 1, \dots, n$, we simplify the notation by writing $A_{n,r}(\theta) := A_n(\theta; \{1, \dots, r\})$, $B_{n,r}(\theta) := B_n(\theta; \{1, \dots, r\})$ and $C_{n,r}(\theta) := C_n(\theta; \{1, \dots, r\})$. With a slight abuse of notation, we use $C_n(\theta)$ for $r = n$ as defined before. Under the enforced assumptions, exchangeability yields

$$\mathbb{P}[A_n(\theta; S)] = \mathbb{P}[A_{n,r}(\theta)], \quad S \in \mathcal{N}_{n,r}$$

and the expression

$$\sum_{S \in \mathcal{N}_{n,r}} \mathbb{P}[A_n(\theta; S)] = \binom{n}{r} \mathbb{P}[A_{n,r}(\theta)] \quad (71)$$

follows since $|\mathcal{N}_{n,r}| = \binom{n}{r}$. Substituting into (70) we obtain the key bound

$$\mathbb{P}[C_n(\theta)^c \cap I(n; \theta)] \leq \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}[A_{n,r}(\theta)]. \quad (72)$$

Consider a scaling $\theta : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \times (0, 1)$ as in the statement of Proposition 11.1. Substitute θ by θ_n by means of this scaling

in the right hand side of (72). The proof of Proposition 11.1 will be completed once we show

$$\lim_{n \rightarrow \infty} \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}[A_{n,r}(\theta_n)] = 0. \quad (73)$$

The means to do so are provided in the next section.

XII. BOUNDING PROBABILITIES

Fix $n = 2, 3, \dots$ and consider $\theta = (K, p)$ with p in $(0, 1)$ and positive integer K such that $K < n$.

A. Bounding the probabilities $\mathbb{P}[B_{n,r}(\theta)]$

The following result will be used to efficiently bound the probability $\mathbb{P}[B_{n,r}(\theta)]$.

Lemma 12.1: For each $r = 2, \dots, n-1$, we have the inequality

$$\begin{aligned} \mathbb{P}[B_{n,r}(\theta) | \Gamma_{n,1}, \dots, \Gamma_{n,r}] \\ \leq (1-p)^{E_{n,r}^*} \cdot u_n(\theta)^{r(n-r)-E_{n,r}^*} \end{aligned} \quad (74)$$

with $u_n(\theta)$ defined by (55) and the rv $E_{n,r}^*$ given by

$$E_{n,r}^* := \sum_{i=r+1}^n \sum_{\ell=1}^r \mathbf{1}[\ell \in \Gamma_{n,i}]. \quad (75)$$

A proof of Lemma 12.1 is available in Appendix B. The rv $E_{n,r}^*$, which appears prominently in (74), has a tail controlled through the following result.

Lemma 12.2: Fix $r = 2, \dots, n-1$. For any t in $(0, 1)$ we have

$$\mathbb{P}[E_{n,r}^* \leq (1-t)rK \cdot \frac{n-r}{n-1}] \leq e^{-\frac{t^2}{2} rK \cdot \frac{n-r}{n-1}}. \quad (76)$$

Proof. Fix $n = 2, 3, \dots$ and consider a positive integer K such that $K < n$. From the facts reported in Section IX, the negative association of the rvs (50) implies that of the rvs $\{\mathbf{1}[\ell \in \Gamma_{n,i}], i = r+1, \dots, n; \ell = 1, \dots, r\}$. We are now in position to apply the Chernoff-Hoeffding bound to the sum (75). We use the bound in the form

$$\mathbb{P}[E_{n,r}^* \leq (1-t)\mathbb{E}[E_{n,r}^*]] \leq e^{-\frac{t^2}{2} \mathbb{E}[E_{n,r}^*]} \quad (77)$$

as given for negatively associated rvs in [10, Thm. 1.1, p. 6]. The conclusion (76) follows upon noting that

$$\mathbb{E}[E_{n,r}^*] = \sum_{i=r+1}^n \sum_{\ell=1}^r \mathbb{P}[\ell \in \Gamma_{n,i}] = r(n-r) \frac{K}{n-1}$$

as we use (31). ■

B. Bounding the probabilities $\mathbb{P}[C_{n,r}(\theta)]$

For each $r = 2, \dots, n$, let $\mathbb{H} \cap \mathbb{G}_r(n; \theta)$ stand for the subgraph $\mathbb{H} \cap \mathbb{G}(n; \theta)(S)$ when $S = \{1, \dots, r\}$. Also let \mathcal{T}_r denote the collection of all spanning trees on the vertex set $\{1, \dots, r\}$.

Lemma 12.3: Fix $r = 2, \dots, n$. For each T in \mathcal{T}_r , we have

$$\mathbb{P}[T \subset \mathbb{H} \cap \mathbb{G}_r(n; \theta)] \leq (p\lambda_n(K))^{r-1} \quad (78)$$

where the notation $T \subset \mathbb{H} \cap \mathbb{G}_r(n; \theta)$ indicates that the tree T is a subgraph spanning $\mathbb{H} \cap \mathbb{G}_r(n; \theta)$.

Since $p\lambda_n(K)$ is the probability of link assignment, the situation is reminiscent to the one found in ER graphs [5] and random key graphs [23] where in each case the bound (78) holds with equality.

Proof. Fix $r = 2, 3, \dots, n$ and pick a tree T in \mathcal{T}_r . Let $\mathcal{E}(T)$ be the set of edges that appear in T . It is plain that $T \subseteq \mathbb{H} \cap \mathbb{G}_r(n; \theta)$ occurs if and only if the set of conditions

$$\begin{aligned} \Sigma_{n,i} \cap \Sigma_{n,j} &\neq \emptyset \\ \text{and} & \\ B_{ij}(p) &= 1 \end{aligned}, \quad \{i, j\} \in \mathcal{E}(T)$$

holds. Therefore, under the enforced independence assumptions, since $|\mathcal{E}(T)| = r - 1$, we get

$$\begin{aligned} \mathbb{P}[T \subset \mathbb{H} \cap \mathbb{G}_r(n; \theta)] &= p^{r-1} \cdot \mathbb{E} \left[\prod_{i,j:\{i,j\} \in \mathcal{E}(T)} \mathbf{1}[\Sigma_{n,i} \cap \Sigma_{n,j} \neq \emptyset] \right] \\ &= p^{r-1} \cdot \mathbb{E} \left[\prod_{i,j:\{i,j\} \in \mathcal{E}(T)} \mathbf{1}[i \in \Gamma_{n,j} \vee j \in \Gamma_{n,i}] \right] \\ &\leq p^{r-1} \cdot \prod_{i,j:\{i,j\} \in \mathcal{E}(T)} \mathbb{P}[i \in \Gamma_{n,j} \vee j \in \Gamma_{n,i}] \end{aligned} \quad (79)$$

by making use of (49) with the negatively associated rvs (46). The desired result (78) is now immediate from (5) and the relation $|\mathcal{E}(T)| = r - 1$. ■

As in ER graphs [5] and random key graphs [23] we have to the following bound.

Lemma 12.4: For each $r = 2, \dots, n$, we have

$$\mathbb{P}[C_{n,r}(\theta)] \leq r^{r-2} (p\lambda_n(K))^{r-1}. \quad (80)$$

Proof. Fix $r = 2, \dots, n$. If $\mathbb{H} \cap \mathbb{G}_r(n; \theta)$ is a connected graph, then it must contain a spanning tree on the vertex set $\{1, \dots, r\}$, and a union bound argument yields

$$\mathbb{P}[C_{n,r}(\theta)] \leq \sum_{T \in \mathcal{T}_r} \mathbb{P}[T \subset \mathbb{H} \cap \mathbb{G}(n; \theta)(S)]. \quad (81)$$

By Cayley's formula [16] there are r^{r-2} trees on r vertices, i.e., $|\mathcal{T}_r| = r^{r-2}$, and (80) follows upon making use of (78). ■

XIII. A PROOF OF PROPOSITION 11.1 (PART II)

Consider a scaling $\theta : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \times (0, 1)$ as in the statement of Proposition 11.1. Pick integers $R \geq 2$ and $n^*(R) \geq 2(R + 1)$ (to be specified in Section XIII-B). On the range $n \geq n^*(R)$ we consider the decomposition

$$\begin{aligned} &\sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}[A_{n,r}(\theta_n)] \\ &= \sum_{r=2}^R \binom{n}{r} \mathbb{P}[A_{n,r}(\theta_n)] + \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}[A_{n,r}(\theta_n)], \end{aligned}$$

and let n go to infinity. The desired convergence (73) will be established if we show

$$\lim_{n \rightarrow \infty} \binom{n}{r} \mathbb{P}[A_{n,r}(\theta_n)] = 0 \quad (82)$$

for each $r = 2, 3, \dots$ and

$$\lim_{n \rightarrow \infty} \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}[A_{n,r}(\theta_n)] = 0. \quad (83)$$

We establish (82) and (83) in turn. Throughout, we make use of the standard bounds

$$\binom{n}{r} \leq \left(\frac{en}{r}\right)^r, \quad r = 1, \dots, n \quad (84)$$

for each $n = 2, 3, \dots$

A. Establishing (82)

Fix $r = 2, 3, \dots$ and consider $n = 2, 3, \dots$ such that $r < n$. Also let $\theta = (K, p)$ with p in $(0, 1)$ and positive integer K such that $K < n$. With (75) in mind, for each $i = 1, \dots, r$, we note that

$$\begin{aligned} \sum_{k=r+1}^n \mathbf{1}[k \in \Gamma_{n,i}] &= \sum_{k=1}^n \mathbf{1}[k \in \Gamma_{n,i}] - \sum_{k=1}^r \mathbf{1}[k \in \Gamma_{n,i}] \\ &= K - \sum_{k=1}^r \mathbf{1}[k \in \Gamma_{n,i}] \end{aligned} \quad (85)$$

since $|\Gamma_{n,i}| = K$. The bounds

$$(K - r)^+ \leq \sum_{k=r+1}^n \mathbf{1}[k \in \Gamma_{n,i}] \leq K$$

follow, whence

$$r(K - r)^+ \leq E_{n,r}^* \leq rK.$$

It is also the case that

$$r(n - r - K)^+ \leq r(n - r) - E_{n,r}^*.$$

Reporting these lower bounds into (74), we get

$$\begin{aligned} &\mathbb{P}[B_{n,r}(\theta) | \Gamma_{n,1}, \dots, \Gamma_{n,r}] \\ &\leq (1 - p)^{r(K-r)^+} \cdot u_n(\theta)^{r(n-r-K)^+} \\ &\leq (1 - p)^{r(K-r)} \cdot u_n(\theta)^{r(n-r-K)} \end{aligned} \quad (86)$$

since $0 < p, u_n(\theta) < 1$. If we set

$$F_{n,r}(\theta) := (1-p)^{(K-r)} \cdot u_n(\theta)^{(n-r-K)},$$

it is now plain that

$$\begin{aligned} \mathbb{P}[A_{n,r}(\theta)] &= \mathbb{E} \left[\mathbf{1}[C_{n,r}(\theta)] \mathbb{P} \left[B_{n,r}(\theta) \middle| \Gamma_{n,1}, \dots, \Gamma_{n,r} \right] \right] \\ &\leq \mathbb{P}[C_{n,r}(\theta)] \cdot F_{n,r}(\theta)^r. \end{aligned} \quad (87)$$

Applying Lemma 12.4 we find

$$\begin{aligned} \binom{n}{r} \mathbb{P}[A_{n,r}(\theta)] &\leq \binom{n}{r} \mathbb{P}[C_{n,r}(\theta)] \cdot F_{n,r}(\theta)^r \\ &\leq \left(\frac{en}{r} \right)^r r^{r-2} (p\lambda_n(K))^{r-1} F_{n,r}(\theta)^r \\ &= \frac{1}{r^2} (en)^r (p\lambda_n(K))^{r-1} F_{n,r}(\theta)^r \end{aligned} \quad (88)$$

as we make use of (84).

We also note that

$$F_{n,r}(\theta) \leq e^{F_{n,r}^*(\theta)} \quad (89)$$

with

$$\begin{aligned} F_{n,r}^*(\theta) &:= (K-r) \log(1-p) - (n-r-K) p \frac{K}{n-1} \\ &= (K-r) \log(1-p) - \left(1 - \frac{K}{n-1} - \frac{r-1}{n-1} \right) pK \\ &= (K-r) \log(1-p) - p \left(K - \frac{K^2}{n-1} \right) + \frac{r-1}{n-1} pK \\ &= K(p + \log(1-p)) - r \log(1-p) \\ &\quad - p \left(2K - \frac{K^2}{n-1} \right) + \frac{r-1}{n-1} pK. \end{aligned} \quad (90)$$

Now, pick any given positive integer $r = 2, 3, \dots$ and consider a scaling $\theta : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \times (0, 1)$ such that $\lim_{n \rightarrow \infty} p_n = p^*$ exists and (10) holds for some $c > \tau(p^*)$. Replace θ by θ_n in (88) according to this scaling. In order to establish (82) it suffices to show that

$$\lim_{n \rightarrow \infty} (en)^r (p_n \lambda_n(K_n))^{r-1} \cdot F_{n,r}(\theta_n)^r = 0. \quad (91)$$

For n sufficiently large, from (12) and (88) we first get

$$\begin{aligned} \binom{n}{r} \mathbb{P}[A_{n,r}(\theta)] &\leq (en)^r (p_n \lambda_n(K_n))^{r-1} \cdot F_{n,r}(\theta_n)^r \\ &= (en)^r \left(c_n \frac{\log n}{n-1} \right)^{r-1} \cdot F_{n,r}(\theta_n)^r \\ &= en \left(ec_n \frac{n}{n-1} \log n \right)^{r-1} \cdot F_{n,r}(\theta_n)^r. \end{aligned} \quad (92)$$

On the other hand, upon making use of the bounds at (38), we find

$$\begin{aligned} F_{n,r}^*(\theta_n) &\leq K_n(p_n + \log(1-p_n)) - r \log(1-p_n) \\ &\quad - p_n \left(2K_n - \frac{K_n^2}{n-1} \right) + \frac{r}{n} p_n K_n \\ &= K_n(p_n + \log(1-p_n)) - r \log(1-p_n) \\ &\quad - c_n \log n + \frac{r}{n} p_n K_n \\ &\leq K_n(p_n + \log(1-p_n)) - c_n \log n \\ &\quad - r \log(1-p_n) + \frac{r}{n} c_n \log n \\ &= p_n K_n \left(1 + \frac{\log(1-p_n)}{p_n} \right) - c_n \log n \\ &\quad - r \log(1-p_n) + \frac{r}{n} c_n \log n \\ &\leq \frac{c_n}{2} \log n \cdot \left(1 + \frac{\log(1-p_n)}{p_n} \right) - c_n \log n \\ &\quad - r \log(1-p_n) + \frac{r}{n} c_n \log n \\ &= -\frac{c_n}{2} \cdot \left(1 - \frac{\log(1-p_n)}{p_n} \right) \log n \\ &\quad - r \log(1-p_n) + \frac{r}{n} c_n \log n. \\ &= \log n \left(-\frac{c_n - \frac{2rp_n}{\log n}}{2} \left(1 - \frac{\log(1-p_n)}{p_n} \right) \right) \\ &\quad - rp_n + \frac{r}{n} c_n \log n \\ &\leq -\frac{\log n}{2} \left(c_n - \frac{2rp_n}{\log n} \right) \left(1 - \frac{\log(1-p_n)}{p_n} \right) \\ &\quad + \frac{r}{n} c_n \log n. \end{aligned} \quad (93)$$

As a result, (90) implies

$$\begin{aligned} n F_{n,r}(\theta_n)^r &\leq n^{1-\frac{r}{2}} (c_n - \frac{2rp_n}{\log n}) \cdot \left(1 - \frac{\log(1-p_n)}{p_n} \right) e^{o(1)}. \end{aligned} \quad (94)$$

Under the enforced assumptions of Theorem 4.2 we get

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 - \frac{r}{2} \left(c_n - \frac{2rp_n}{\log n} \right) \cdot \left(1 - \frac{\log(1-p_n)}{p_n} \right) \right) \\ = 1 - r \frac{c}{2} \cdot \left(1 - \frac{\log(1-p^*)}{p^*} \right) \\ = 1 - r \frac{c}{\tau(p^*)} < 0, \end{aligned} \quad (95)$$

and the desired conclusion (91) follows upon making use of the inequalities (92) and (94).

B. Establishing (83)

Fix $n = 2, 3, \dots$ and consider $\theta = (K, p)$ with p in $(0, 1)$, and positive integer K such that $K < n$.

Pick $r = 1, 2, \dots, n-1$. By Lemma 12.1 we conclude that

$$\mathbb{P} \left[B_{n,r}(\theta) \middle| \Gamma_{n,1}, \dots, \Gamma_{n,r} \right] \leq (1-p)^{E_{n,r}^*} \quad (96)$$

since $0 < u_n(\theta) < 1$, and preconditioning arguments similar to the ones leading to (87) yield

$$\mathbb{P}[A_{n,r}(\theta)] \leq \mathbb{E} \left[\mathbf{1}[C_{n,r}(\theta)] (1-p)^{E_{n,r}^*} \right].$$

The event $C_{n,r}(\theta)$ depends only on $\Gamma_{n,1}, \dots, \Gamma_{n,r}$ whereas $E_{n,r}^*$ is determined solely by $\Gamma_{n,r+1}, \dots, \Gamma_{n,n}$. Thus, the event $C_{n,r}(\theta)$ is independent of the rv $(1-p)^{E_{n,r}^*}$ under the enforced assumptions, whence

$$\mathbb{P}[A_{n,r}(\theta)] \leq \mathbb{P}[C_{n,r}(\theta)] \mathbb{E} \left[(1-p)^{E_{n,r}^*} \right]. \quad (97)$$

Pick t arbitrary in $(0, 1)$ and recall Lemma 12.2. A simple decomposition argument shows that

$$\begin{aligned} & \mathbb{E} \left[(1-p)^{E_{n,r}^*} \right] \\ & \leq \mathbb{E} \left[(1-p)^{E_{n,r}^*} \mathbf{1} \left[E_{n,r}^* > (1-t)rK \cdot \frac{n-r}{n-1} \right] \right] \\ & \quad + \mathbb{P} \left[E_{n,r}^* \leq (1-t)rK \cdot \frac{n-r}{n-1} \right] \\ & \leq (1-p)^{(1-t)rK \cdot \frac{n-r}{n-1}} + e^{-\frac{t^2}{2}rK \cdot \frac{n-r}{n-1}} \\ & \leq e^{-(1-t)rpK \cdot \frac{n-r}{n-1}} + e^{-\frac{t^2}{2}rK \cdot \frac{n-r}{n-1}} \\ & \leq e^{-(1-t)rpK \cdot \frac{n-r}{n-1}} + e^{-\frac{t^2}{2}rpK \cdot \frac{n-r}{n-1}}. \end{aligned}$$

Therefore, whenever $r = 2, 3, \dots, \lfloor \frac{n}{2} \rfloor$, we have

$$\mathbb{E} \left[(1-p)^{E_{n,r}^*} \right] \leq e^{-\frac{1-t}{2}rpK} + e^{-\frac{t^2}{4}rpK} \quad (98)$$

since on that range we have

$$\frac{n-r}{n-1} \geq \frac{n/2}{n-1} \geq \frac{1}{2}.$$

Now consider a scaling $\theta : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \times (0, 1)$ such that $\lim_{n \rightarrow \infty} p_n = p^*$ exists and (10) holds for some $c > \tau(p^*)$. Replace θ by θ_n in both (97) and (98) according to this scaling and use the bound of Lemma 12.4 in the resulting inequalities. Pick an integer $R \geq 2$ (to be further specified shortly) and for $n \geq 2(R+1)$ note that

$$\begin{aligned} & \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}[A_{n,r}(\theta_n)] \\ & \leq \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} r^{r-2} (p_n \lambda_n(K_n))^{r-1} e^{-\frac{1-t}{2}rp_n K_n} \\ & \quad + \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} r^{r-2} (p_n \lambda_n(K_n))^{r-1} e^{-\frac{t^2}{4}rp_n K_n} \\ & \leq \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} en \left(ec_n \frac{n}{n-1} \log n \right)^{r-1} e^{-\frac{1-t}{2}rp_n K_n} \\ & \quad + \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} en \left(ec_n \frac{n}{n-1} \log n \right)^{r-1} e^{-\frac{t^2}{4}rp_n K_n} \end{aligned}$$

by the same arguments as the ones leading to (92). Upon invoking the lower bound in (38) we now conclude for all sufficiently large $n > 2(R+1)$ that

$$\begin{aligned} & \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}[A_{n,r}(\theta_n)] \\ & \leq \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} en \left(ec_n \frac{n}{n-1} \log n \right)^r e^{-\frac{1-t}{4}rc_n \log n} \\ & \quad + \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} en \left(ec_n \frac{n}{n-1} \log n \right)^r e^{-\frac{t^2}{8}rc_n \log n} \\ & \leq \sum_{r=R+1}^{\infty} en \left(ec_n \frac{n}{n-1} \log n \cdot n^{-\frac{1-t}{4}c_n} \right)^r \\ & \quad + \sum_{r=R+1}^{\infty} en \left(ec_n \frac{n}{n-1} \log n \cdot n^{-\frac{t^2}{8}c_n} \right)^r. \end{aligned}$$

Furthermore, for all sufficiently large $n \geq 2(R+1)$ it also the case that

$$ec_n \frac{n}{n-1} \log n \cdot \max \left(n^{-\frac{1-t}{4}c_n}, n^{-\frac{t^2}{8}c_n} \right) < 1 \quad (99)$$

and the two infinite series converge. Let $n^*(R)$ denote any integer larger than $2(R+1)$ such that (99) holds for all $n \geq n^*(R)$. On that range, by our earlier discussion we get

$$\sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}[A_{n,r}(\theta_n)] \leq e \left(ec_n \frac{n}{n-1} \log n \right)^{R+1} (\dots)$$

with

$$\begin{aligned} \dots & := \frac{n^{1-\frac{1-t}{4}c_n(R+1)}}{1 - ec_n \frac{n}{n-1} \log n \cdot n^{-\frac{1-t}{4}c_n}} \\ & \quad + \frac{n^{1-\frac{t^2}{8}c_n(R+1)}}{1 - ec_n \frac{n}{n-1} \log n \cdot n^{-\frac{t^2}{8}c_n}}. \end{aligned}$$

Finally, let n go to infinity in this last expression: The desired conclusion (83) follows whenever the conditions $(1-t)c(R+1) > 4$ and $c(R+1)t^2 > 8$ are satisfied. This can be achieved by taking R so that

$$R+1 > \max \left(\frac{4}{c(1-t)}, \frac{8}{ct^2} \right).$$

This is always feasible for any given t in $(0, 1)$ by taking R sufficiently large. ■

APPENDIX A

A PROOF OF PROPOSITION 10.1

The basis for deriving (58) lies in the observation that nodes 1 and 2 are both isolated in $\mathbb{H} \cap \mathbb{G}(n; \theta)$ if and only if each edge in $\mathbb{H}(n; K)$ incident to one of these nodes is *not* present in $\mathbb{G}(n; p)$. Thus, $\chi_{n,1}(\theta) = \chi_{n,2}(\theta) = 1$ if and only if both sets of conditions

$$B_{1j}(p) = 0 \quad \text{if} \quad \Sigma_{n,1} \cap \Sigma_{n,j} \neq \emptyset, \quad j \in \mathcal{N}_{-1}$$

and

$$B_{2k}(p) = 0 \quad \text{if} \quad \Sigma_{n,2} \cap \Sigma_{n,k} \neq \emptyset, \quad k \in \mathcal{N}_{-2}$$

hold.

To formalize this observation, we introduce the random sets $N_{n,1}(\theta)$ and $N_{n,2}(\theta)$ defined by

$$N_{n,1}(\theta) := \{j = 3, \dots, n : j \in \Gamma_{n,1} \vee 1 \in \Gamma_{n,j}\} \quad (100)$$

and

$$N_{n,2}(\theta) := \{k = 3, \dots, n : k \in \Gamma_{n,2} \vee 2 \in \Gamma_{n,k}\}. \quad (101)$$

Thus, node j in $N_{n,1}(\theta)$ is neither node 1 nor node 2, and is K-adjacent to node 1. Similarly, node k in $N_{n,2}(\theta)$ is neither node 1 nor node 2, and is K-adjacent to node 2. Let $Z_n(\theta)$ denote the total number of edges in $\mathbb{H}(n; K)$ which are incident to either node 1 or node 2. It is plain that

$$\begin{aligned} Z_n(\theta) &= |N_{n,1}(\theta)| + |N_{n,2}(\theta)| \\ &\quad + \mathbf{1}[2 \in \Gamma_{n,1} \vee 1 \in \Gamma_{n,2}] \end{aligned} \quad (102)$$

with the last term accounting for the possibility that nodes 1 and 2 are K-adjacent. By conditioning on the rvs $\Gamma_{n,1}, \dots, \Gamma_{n,n}$, we readily conclude that

$$\mathbb{E}[\chi_{n,1}(\theta)\chi_{n,2}(\theta)] = \mathbb{E}[(1-p)^{Z_n(\theta)}] \quad (103)$$

under the enforced independence of the collections of rvs $\{\Gamma_{n,1}, \dots, \Gamma_{n,n}\}$ and $\{B_{ij}(p), 1 \leq i < j \leq n\}$.

To proceed we need to assess the various contributions to $Z_n(\theta)$: Using (1) we find

$$\begin{aligned} |N_{n,1}(\theta)| &= \sum_{j=3}^n \mathbf{1}[j \in \Gamma_{n,1} \vee 1 \in \Gamma_{n,j}] \\ &= \sum_{j=3}^n \mathbf{1}[j \in \Gamma_{n,1}] + \sum_{j=3}^n \mathbf{1}[1 \in \Gamma_{n,j}] \\ &\quad - \sum_{j=3}^n \mathbf{1}[j \in \Gamma_{n,1}, 1 \in \Gamma_{n,j}] \\ &= \sum_{j=3}^n \mathbf{1}[j \in \Gamma_{n,1}] + \sum_{j=3}^n \mathbf{1}[j \notin \Gamma_{n,1}, 1 \in \Gamma_{n,j}] \\ &= K - \mathbf{1}[2 \in \Gamma_{n,1}] \\ &\quad + \sum_{j=3}^n \mathbf{1}[j \notin \Gamma_{n,1}, 1 \in \Gamma_{n,j}] \end{aligned} \quad (104)$$

where the last step used the fact $|\Gamma_{n,1}| = K$. Similar arguments show that

$$\begin{aligned} |N_{n,2}(\theta)| &= \sum_{k=3}^n \mathbf{1}[k \in \Gamma_{n,2} \vee 2 \in \Gamma_{n,k}] \\ &= K - \mathbf{1}[1 \in \Gamma_{n,2}] \\ &\quad + \sum_{k=3}^n \mathbf{1}[k \notin \Gamma_{n,2}, 2 \in \Gamma_{n,k}]. \end{aligned} \quad (105)$$

As a result, from the definition of $Z_n(\theta)$ we get

$$Z_n(\theta) = 2K - \mathbf{1}[2 \in \Gamma_{n,1}, 1 \in \Gamma_{n,2}] + Z_n^*(\theta) \quad (106)$$

upon using (1) one more time, where

$$\begin{aligned} Z_n^*(\theta) &:= \sum_{j=3}^n \mathbf{1}[j \notin \Gamma_{n,1}, 1 \in \Gamma_{n,j}] \\ &\quad + \sum_{j=3}^n \mathbf{1}[j \notin \Gamma_{n,2}, 2 \in \Gamma_{n,j}]. \end{aligned} \quad (107)$$

In order to evaluate the expression (103), we first compute the conditional expectation

$$\mathbb{E}[(1-p)^{Z_n(\theta)} | \Gamma_{n,1}, \Gamma_{n,2}]. \quad (108)$$

From (106) we see that this quantity can be evaluated as the product of the two terms

$$(1-p)^{2K - (\mathbf{1}[2 \in \Gamma_{n,1}, 1 \in \Gamma_{n,2}])} \quad (109)$$

and

$$\mathbb{E}[(1-p)^{Z_n^*(\theta)} | \Gamma_{n,1}, \Gamma_{n,2}]. \quad (110)$$

To evaluate this last conditional expectation, for each $j = 3, \dots, n$, we set

$$\begin{aligned} V_{n,j}(\theta; S, T) &:= \mathbb{E}[(1-p)^{\mathbf{1}[j \notin S, 1 \in \Gamma_{n,j}] + \mathbf{1}[j \notin T, 2 \in \Gamma_{n,j}]}] \end{aligned}$$

with S and T subsets of \mathcal{N} , each being of size K . It is straightforward to check that

$$\begin{aligned} V_{n,j}(\theta; S, T) &= \mathbf{1}[j \notin S] \mathbf{1}[j \notin T] \mathbb{E}[(1-p)^{\mathbf{1}[1 \in \Gamma_{n,j}] + \mathbf{1}[2 \in \Gamma_{n,j}]}] \\ &\quad + \mathbf{1}[j \notin S] \mathbf{1}[j \in T] \mathbb{E}[(1-p)^{\mathbf{1}[1 \in \Gamma_{n,j}]}] \\ &\quad + \mathbf{1}[j \in S] \mathbf{1}[j \notin T] \mathbb{E}[(1-p)^{\mathbf{1}[2 \in \Gamma_{n,j}]}] \\ &\quad + \mathbf{1}[j \in S] \mathbf{1}[j \in T]. \end{aligned}$$

Then, with the notation introduced earlier in Section IX, we can write

$$\begin{aligned} V_{n,j}(\theta; S, T) &= \mathbf{1}[j \notin S] \mathbf{1}[j \notin T] b_n(\theta) \\ &\quad + (\mathbf{1}[j \notin S] \mathbf{1}[j \in T] + \mathbf{1}[j \in S] \mathbf{1}[j \notin T]) u_n(\theta) \\ &\quad + \mathbf{1}[j \in S] \mathbf{1}[j \in T]. \end{aligned}$$

Next, the two rvs $\Gamma_{n,1}$ and $\Gamma_{n,2}$ being jointly independent of the rvs $\Gamma_{n,3}, \dots, \Gamma_{n,n}$, we find

$$\begin{aligned} \mathbb{E}[(1-p)^{Z_n^*(\theta)} | \Gamma_{n,1}, \Gamma_{n,2}] &= \prod_{j=3}^n V_{n,j}(\theta; \Gamma_{n,1}, \Gamma_{n,2}) \\ &= b_n(\theta)^{B_n(\theta)} \cdot u_n(\theta)^{U_n(\theta)} \end{aligned} \quad (111)$$

where the rvs $B_n(\theta)$ and $U_n(\theta)$ are given by (56) and (57), respectively. Therefore, since

$$\mathbb{E}[(1-p)^{Z_n(\theta)}] = \mathbb{E}[\mathbb{E}[(1-p)^{Z_n(\theta)} | \Gamma_{n,1}, \Gamma_{n,2}]]$$

by a standard preconditioning argument, we get the expression (58) upon writing (108) as the product of the quantities (109) and (110), and using (111). ■

APPENDIX B A PROOF OF LEMMA 12.1

The defining conditions for $B_{n,r}(\theta)$ lead to the representation

$$B_{n,r}(\theta) = \cap_{i=1}^r \cap_{k=r+1}^n E_{n,ik}(\theta)$$

where we have set

$$E_{n,ik}(\theta) := ([k \notin \Gamma_{n,i}] \cap [i \notin \Gamma_{n,k}]) \cup [B_{ik}(p) = 0]$$

with $i = 1, \dots, r$ and $k = r+1, \dots, n$. In terms of indicator functions, with the help of (1) this definition reads

$$\begin{aligned} \mathbf{1}[E_{n,ik}(\theta)] &= \mathbf{1}[k \notin \Gamma_{n,i}] \mathbf{1}[i \notin \Gamma_{n,k}] + (1 - B_{ik}(p)) \\ &\quad - \mathbf{1}[k \notin \Gamma_{n,i}] \mathbf{1}[i \notin \Gamma_{n,k}] (1 - B_{ik}(p)) \\ &= (1 - B_{ik}(p)) + \mathbf{1}[k \notin \Gamma_{n,i}] \mathbf{1}[i \notin \Gamma_{n,k}] B_{ik}(p). \end{aligned}$$

Therefore, under the enforced independence assumptions,

$$\begin{aligned} \mathbb{P}[B_{n,r}(\theta) | \Gamma_{n,1}, \dots, \Gamma_{n,n}] \\ = \mathbb{E} \left[\prod_{i=1}^r \prod_{k=r+1}^n W(\mathbf{1}[k \notin \Gamma_{n,i}] \mathbf{1}[i \notin \Gamma_{n,k}]; p) \right] \end{aligned}$$

where

$$W(x; p) = 1 - p + px, \quad x \in \mathbb{R}.$$

Since $W(x, p) = (1 - p)^{1-x}$ for $x = 0, 1$, we obtain

$$\begin{aligned} \mathbb{P}[B_{n,r}(\theta) | \Gamma_{n,1}, \dots, \Gamma_{n,n}] \\ = \mathbb{E} \left[\prod_{i=1}^r \prod_{k=r+1}^n (1 - p)^{1 - \mathbf{1}[k \notin \Gamma_{n,i}] \mathbf{1}[i \notin \Gamma_{n,k}]} \right], \end{aligned}$$

and it is now plain that

$$\begin{aligned} \mathbb{P}[B_{n,r}(\theta) | \Gamma_{n,1}, \dots, \Gamma_{n,r}] \\ = (1 - p)^{r(n-r)} G_{n,r}(\Gamma_{n,1}, \dots, \Gamma_{n,r}; \theta) \end{aligned}$$

where we have set

$$\begin{aligned} G_{n,r}(S_1, \dots, S_r; \theta) \\ = \mathbb{E} \left[\prod_{i=1}^r \prod_{k=r+1}^n (1 - p)^{-\mathbf{1}[k \notin S_i] \mathbf{1}[i \notin \Gamma_{n,k}]} \right] \end{aligned}$$

with S_1, \dots, S_r subsets of \mathcal{N} , each of size K .

Next, we find

$$\begin{aligned} G_{n,r}(S_1, \dots, S_r; \theta) \\ = \mathbb{E} \left[\prod_{k=r+1}^n \prod_{i=1}^r (1 - p)^{-\mathbf{1}[k \notin S_i] \mathbf{1}[i \notin \Gamma_{n,k}]} \right] \\ = \mathbb{E} \left[\prod_{k=r+1}^n (1 - p)^{-\sum_{i=1}^r \mathbf{1}[k \notin S_i] \mathbf{1}[i \notin \Gamma_{n,k}]} \right] \\ = \prod_{k=r+1}^n \mathbb{E} \left[(1 - p)^{-\sum_{i=1}^r \mathbf{1}[k \notin S_i] \mathbf{1}[i \notin \Gamma_{n,k}]} \right] \end{aligned}$$

as we again use the enforced independence assumptions. Fix $k = r+1, \dots, n$ and note that

$$\begin{aligned} \mathbb{E} \left[(1 - p)^{-\sum_{i=1}^r \mathbf{1}[k \notin S_i] \mathbf{1}[i \notin \Gamma_{n,k}]} \right] \\ = \mathbb{E} \left[\prod_{i=1}^r \left((1 - p)^{-\mathbf{1}[k \notin S_i]} \right)^{\mathbf{1}[i \notin \Gamma_{n,k}]} \right] \\ \leq \prod_{i=1}^r \mathbb{E} \left[\left((1 - p)^{-\mathbf{1}[k \notin S_i]} \right)^{\mathbf{1}[i \notin \Gamma_{n,k}]} \right] \quad (112) \\ = \prod_{i=1}^r \mathbb{E} \left[(1 - p)^{-\mathbf{1}[i \notin \Gamma_{n,k}]} \right]^{\mathbf{1}[k \notin S_i]} \end{aligned}$$

where (112) follows from the negative association of the rvs (45) – Use (48) and note that

$$(1 - p)^{-\mathbf{1}[k \notin S_i]} \geq 1, \quad i = 1, \dots, r.$$

Next we observe that for each $i = 1, \dots, r$, we have

$$\begin{aligned} \mathbb{E} \left[(1 - p)^{-\mathbf{1}[i \notin \Gamma_{n,k}]} \right] \\ = (1 - p)^{-1} \mathbb{P}[i \notin \Gamma_{n,k}] + \mathbb{P}[i \in \Gamma_{n,k}] \\ = (1 - p)^{-1} \left(1 - \frac{K}{n-1} \right) + \frac{K}{n-1} \\ = \frac{u_n(\theta)}{1 - p} \end{aligned}$$

whence

$$\prod_{i=1}^r \mathbb{E} \left[(1 - p)^{-\mathbf{1}[i \notin \Gamma_{n,k}]} \right]^{\mathbf{1}[k \notin S_i]} = \left(\frac{u_n(\theta)}{1 - p} \right)^{\sum_{i=1}^r \mathbf{1}[k \notin S_i]}.$$

Combining these observations readily yields

$$\begin{aligned} G_{n,r}(S_1, \dots, S_r; \theta) \\ \leq \prod_{k=r+1}^n \left(\frac{u_n(\theta)}{1 - p} \right)^{\sum_{i=1}^r \mathbf{1}[k \notin S_i]} \\ = \left(\frac{u_n(\theta)}{1 - p} \right)^{\sum_{i=1}^r \sum_{k=r+1}^n \mathbf{1}[k \notin S_i]}. \end{aligned}$$

We finally obtain

$$\begin{aligned} \mathbb{P}[B_{n,r}(\theta) | \Gamma_{n,1}, \dots, \Gamma_{n,r}] \\ \leq (1 - p)^{r(n-r)} \left(\frac{u_n(\theta)}{1 - p} \right)^{\sum_{i=1}^r \sum_{k=r+1}^n \mathbf{1}[k \notin \Gamma_{n,i}]} \end{aligned}$$

and the desired conclusion (74) follows. ■

ACKNOWLEDGMENT

This work was supported by NSF Grant CCF-07290.

REFERENCES

- [1] I. F. Akyildiz, Y. Sankarsubramaniam, W. Su and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks* **38**, pp. 393-422.
- [2] N. P. Anthapadmanabhan and A. M. Makowski, "On the absence of isolated nodes in wireless ad-hoc networks with unreliable links - A curious gap," Proceedings of IEEE Infocom 2010, San Diego (CA), March 2010.
- [3] S.R. Blackburn and S. Gerke, "Connectivity of the uniform random intersection graph," *Discrete Mathematics* **309** (2009), pp. 5130-5140.
- [4] M. Bloznelis, J. Jaworski and K. Rybarczyk, "Component evolution in a secure wireless sensor network," *Networks* **53** (2009), pp. 19-26.
- [5] B. Bollobás, *Random Graphs*, Second Edition, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge (UK), 2001.
- [6] S. A. Çamtepe and B. Yener, "Key Distribution Mechanisms for Wireless Sensor Networks: a Survey," Technical Report TR-05-07, Computer Science Department, Rensselaer Polytechnic Institute, Troy (NY), March 2005.
- [7] H. Chan, A. Perrig and D. Song, "Random key predistribution schemes for sensor networks," Proceedings of the 2003 IEEE Symposium on Research in Security and Privacy (SP 2003), Oakland (CA), May 2003, pp. 197-213.
- [8] R. Di Pietro, L.V. Mancini, A. Mei, A. Panconesi and J. Radhakrishnan, "Redoubtable sensor networks," *ACM Transactions on Information Systems Security TISSEC* **11** (2008), pp. 1-22.
- [9] W. Du, J. Deng, Y.S. Han and P.K. Varshney, "A pairwise key pre-distribution scheme for wireless sensor networks," Proceedings of the 10th ACM Conference on Computer and Communications Security (CCS 2003), Washington (DC), October 2003, pp. 42-51.
- [10] D. Dubhashi and A. Panconesi, *Concentration of Measure for the Analysis of Randomized Algorithms*, Cambridge University Press, New York (NY), 2009.
- [11] L. Eschenauer and V.D. Gligor, "A key-management scheme for distributed sensor networks," Proceedings of the ACM Conference on Computer and Communications Security (CSS 2002), Washington (DC), November 2002, pp. 41-47.
- [12] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks, Chapter in *Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming*, Edited by W.M. McEneaney, G. Yin and Q. Zhang, Birkhauser, Boston (MA), 1998.
- [13] J. Hwang and Y. Kim, "Revisiting random key pre-distribution schemes for wireless sensor networks," Proceedings of the Second ACM Workshop on Security of Ad Hoc And Sensor Networks (SASN 2004), Washington (DC), October 2004.
- [14] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, 2000.
- [15] K. Joag-Dev and F. Proschan, "Negative association of random variables, with applications," *The Annals of Statistics* **11** (1983), pp. 266-295.
- [16] G.E. Martin, *Counting: The Art of Enumerative Combinatorics*, Springer Verlag New York, 2001.
- [17] A. Mei, A. Panconesi and J. Radhakrishnan, "Unassailable sensor networks," Proceedings of the 4th International Conference on Security and Privacy in Communication Networks (SecureComm), Istanbul (Turkey), September 2008.
- [18] K. Rybarczyk, "Diameter, connectivity and phase transition of the uniform random intersection graph," Submitted to *Discrete Mathematics*, July 2009.
- [19] M.D. Penrose, *Random Geometric Graphs*, Oxford Studies in Probability **5**, Oxford University Press, New York (NY), 2003.
- [20] A. Perrig, J. Stankovic and D. Wagner, "Security in wireless sensor networks," *Communications of the ACM* **47** (2004), pp. 53-57.
- [21] D.-M. Sun and B. He, "Review of key management mechanisms in wireless sensor networks," *Acta Automatica Sinica* **12** (2006), pp. 900-906.
- [22] O. Yağan and A.M. Makowski, "On the random graph induced by a random key predistribution scheme under full visibility," Proceedings of the IEEE International Symposium on Information Theory (ISIT 2008), Toronto (ON), June 2008.
- [23] O. Yağan and A. M. Makowski, "Connectivity results for random key graphs," Proceedings of the IEEE International Symposium on Information Theory (ISIT 2009), Seoul (Korea), June 2009.
- [24] O. Yağan and A.M. Makowski, "Zero-one laws for connectivity in random key graphs," Available online at arXiv:0908.3644v1 [math.CO], August 2009. Earlier draft available online at <http://hdl.handle.net/1903/8716>, January 2009.
- [25] O. Yağan and A. M. Makowski, "On random graphs associated with a pairwise key distribution scheme for wireless sensor networks (Extended version)," submitted for inclusion in the program of IEEE Infocom 2011, Shanghai (PRC), April 2011. Available online at <http://hdl.handle.net/1903/10601>.
- [26] O. Yağan and A. M. Makowski, "On the gradual deployment of random pairwise key distribution schemes," submitted for inclusion in the program of IEEE Infocom 2011, Shanghai (PRC), April 2011. Available online at <http://hdl.handle.net/1903/10604>.
- [27] O. Yağan and A. M. Makowski, "Designing securely connected wireless sensor networks in the presence of unreliable links," submitted for inclusion in the program of ICC 2011, Tokyo (Japan), June 2011.
- [28] C.W. Yi, P.J. Wan, K.W. Lin and C.H. Huang, "Asymptotic distribution of the number of isolated nodes in wireless ad hoc networks with unreliable nodes and links," Proceedings of IEEE Globecom 2006, San Francisco (CA), November 2006.